

The case for the JASONS

Two of the most secretive research operations in the United States have chosen a strange time to embark on an unseemly and unnecessary public row, potentially undermining the advice available to the administration on national defence.

The Defense Advanced Research Projects Agency (DARPA) — the advanced research wing of the Pentagon — is in the process of severing its links with the JASONS, an esteemed élite group of scientific advisers that has independently reviewed classified research for decades (see page 353).

The news of this breach, as the United States stands at the brink of unknown military actions in its proclaimed war against terrorism, could hardly come at a more inauspicious time. So why has the break occurred? And why has it happened so publicly? According to DARPA officials, the agency has been involved in protracted and ultimately unsuccessful contract negotiations with the Miter Corporation, which administers the JASONS.

At the same time, however, a spokeswoman for the research agency has been ruthlessly shredding the JASON group's considerable reputation. The allegation is that the group, whose unpublished list of members includes around 50 of the most distinguished US scientists from all disciplines, is unfashionably weighted towards physicists and other physical scientists. Modern warfare, according to this improbable spin, requires a greater infusion of an entirely new generation of skills in information technology, biowarfare, and so forth, that the JASONS don't possess.

The facts are rather different. The group's membership reflects a wide range of technical expertise and experience. But they are academics, not business people or consultants. The director of DARPA, Tony Tether, appears to have lost patience with the JASONS after they

resisted his attempt to foist three new members of his own choosing onto the group. Relations between DARPA and the JASONS appear to have deteriorated when they refused to accept what they regarded as a downgrading of their status to that of many other panels of government-appointed yes-persons. The self-appointed group has never exactly been a model of democracy — but it has possessed a genuine independence that too many advisory groups in other spheres conspicuously lack.

Why the disagreement has surfaced in the media is another matter. The JASONS have little apparent interest in making it public. They already have ample — some would say excessive — connections with people in power in both the administration and the Congress, and must have fancied their chances of fighting and winning the argument behind closed doors. It is possible, instead, that news of the dispute came to National Public Radio, which broke the story, by way of someone, perhaps not a million miles from the office of the DARPA director, who is seeking to discredit Tether.

At a time when emotion and money are flowing freely through the Pentagon, it seems especially important to have a group of objective scientists who owe their appointments to no one and who can assess the most technically challenging problems. The JASONS may be somewhat old-fashioned in their approach, and they may indeed be in need of new blood. But they have a strong track record in providing the unflinching advice that government officials don't always want to hear, and their work should be allowed to continue. ■

Rights, wrongs and ignorance

Whether or not animals have 'rights', we should learn more about their capacity for suffering.

In Germany, the right of freedom to research is enshrined in the nation's constitution. But that may soon have to be balanced against a new constitutional right of animals to be treated as fellow creatures, and sheltered from avoidable pain. Not surprisingly, biomedical researchers fear that their work will be mired in legal challenges.

The latest moves in Germany are the product of political circumstances (see page 355). But attempts to give animal rights a legal foundation are quietly gathering momentum worldwide. Three years ago, New Zealand's parliament considered — and ultimately rejected — a plan to extend basic human rights to the great apes. And at a growing number of law schools in the United States, courses in animal law are popular.

The book *Rattling the Cage: Towards Legal Rights for Animals* (Profile Books/Perseus; 2000) usually features on the reading list for such courses. In this volume, Steven Wise, former president of the Animal Legal Defense Fund, argues that the great apes' cognitive abilities and self-awareness demand that they be granted rights, including that of "bodily integrity". Invasive or harmful biomedical experimentation on these animals should, in Wise's view, be ruled out.

His follow-up, *Drawing the Line: Science and the Case for Animal Rights* (Perseus; 2002), will appear in June. Wise now argues that dolphins belong in the same category as the great apes. For many other species, he asserts that our ignorance of their cognitive abilities is

such that we cannot assess whether they should be afforded rights. In such cases, Wise argues that the precautionary principle should be applied, erring on the side of assigning at least minimal legal rights.

Wise's arguments should be challenged on both legal and scientific grounds. Some commentators have already countered that 'rights' are only created by beings capable of asserting them — therefore very young children, and animals, are properly accorded protection, not rights (see *Nature* **406**, 675–676; 2000). No doubt some cognitive scientists will take issue with Wise's interpretation of the scientific evidence. Many people — and *Nature* — would also apply the precautionary principle rather differently, factoring in the immense benefits of biomedical experimentation for human health.

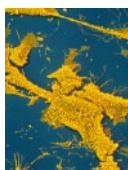
Nevertheless, most experts would agree that we have barely started to understand animal cognition. Even our knowledge of animal welfare is still rudimentary. We can measure levels of hormones that correlate with stress in people. But is a rat with high levels of corticosteroids suffering? We just don't know.

Given the passions raised by animal experimentation, and the importance of biomedical research to human health, the science of animal suffering and cognition should be given a higher priority. We owe it to ourselves, as much as to our fellow creatures, not simply to leave the lawyers to battle it out. ■





Pain killers
Animals set to gain constitutional rights in Germany
p355



Cancer trial
Search for nerve tumours fails its screen test
p356



Twist of fate
Bioinformatics beacon bites the dust
p357



On the rocks
Satellite captures the death throes of an ice shelf
p358

Covert science counsel cut off as nomination row piques Pentagon

Geoff Brumfiel, Washington

An exclusive club of US scientists that advises the government on military matters has lost much of its funding after refusing to accept orders over whom it should take as a member.

The funding was cut after the group — known as the JASONs — rebuffed three nominees put forward by the Defense Advanced Research Projects Agency (DARPA), a division of the Pentagon devoted to innovative research and which provides about half of the think tank's money. Some JASON members say that they saw the nominations as an attempt to make political appointments.

The JASONs, a group of about 50 scientists, have met behind closed doors for the past 40 years to review the Pentagon's toughest technical problems. The team, which is primarily composed of prominent university researchers, has played an important role in the development of modern submarine warfare, for example, and in assuring the government that nuclear weapons can be safely maintained in the absence of testing.

The nomination problem began to unfold last summer, when Tony Tether, who was appointed as DARPA's director by President George Bush, suggested three new members intended to provide the JASONs with expertise in Internet-related technologies. The JASONs were not impressed with the nominees — two Silicon Valley executives and an engineer — and, after reviewing their credentials and willingness to participate, told DARPA that the group would not accept them.

In negotiations that followed, sources familiar with the situation say, the JASONs offered to let the new members join their meetings on an ad-hoc basis or, failing that, proposed that DARPA nominate alternative candidates. DARPA refused and allowed its contract with the JASONs to expire this year.

DARPA officials say that the new members would have brought fresh blood into the JASONs and that without them, the organization has become obsolete. "The JASONs were valuable during the cold war," says a spokeswoman for the agency, "but after it ended, they didn't move towards new fields



Two research operations behind the military prowess of the United States are at daggers drawn.

such as information technology." Without the new appointments, the JASONs would be too heavily oriented towards physical sciences to be of help to DARPA, she says.

"The DARPA statement is just wrong," says Steven Koonin, provost of the California Institute of Technology in Pasadena and current head of the JASONs. Koonin concedes that more than half of the JASONs are physicists, but points out that the group also includes prominent biologists, computer scientists and engineers. Over the past few years, he says, the topics that the group has researched include climate change and biowarfare. "It is amazing to me, given the heightened importance of national security, that we've been sidelined," he says.

Steven Block, a biophysicist at Stanford University and JASON member, and other members of the group who declined to be identified, allege that DARPA's insistence on its three candidates was an attempt to make political appointments to a panel that strongly values its independence. "There seems to be an attempt to make the JASONs into a political patronage job," Block says.

Others close to the dispute think that personal differences led to the split. Conflicts between the JASONs and DARPA officials have erupted before. "This should have been talked out, but somehow it became a matter of principle, and then it got out of hand," a source close to the JASONs says.

The JASONs are now looking for a prime sponsor to replace DARPA. They are believed to be close to finding an alternative within the Pentagon, although they may also talk to other agencies, such as the Central Intelligence Agency and the Department of Energy.

The dispute between the two groups has worried observers inside and outside the US government. "We think it's a mistake," says Steven Aftergood of the Federation of American Scientists, an organization that monitors issues relating to science and defence. Although Aftergood admits to being frustrated in the past with the veil of secrecy that surrounds the JASONs' work, he believes that they provide the Pentagon with much-needed independent advice. "They provide a useful reality check for programmes that otherwise might be running on enthusiasm," he says. ■

Biologists question adult stem-cell versatility

Natalie DeWitt and Jonathan Knight

A fresh debate has broken out over the importance of adult stem cells, in the wake of new evidence that their potential usefulness in medicine has been exaggerated.

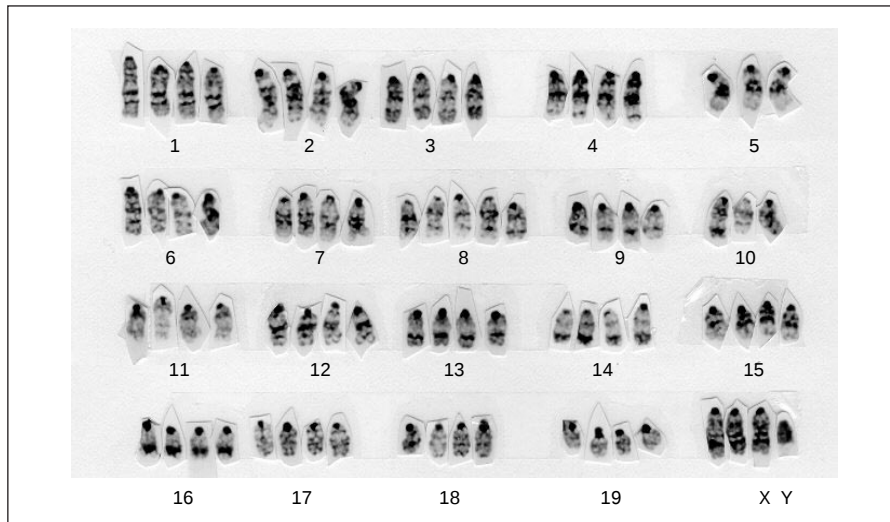
Many researchers believe that embryonic stem cells, which can develop into any type of cell in the body, could revolutionize treatments for conditions ranging from heart disease to Parkinson's disease. Those opposed to such research, because it involves killing human embryos, have touted adult stem cells as an alternative. And over the past two years, evidence has mounted that adult cells may be almost as malleable as embryonic cells.

But two papers published on *Nature's* website on 13 March (DOIs: 10.1038/nature729; 10.1038/nature730) raised doubts about the validity of those results, according to researchers attending a scientific conference on stem cells in Keystone, Colorado.

Stem cells in the adult body have limited options. Certain stem cells from bone marrow, for example, can form only blood. But several reports in the past two years have hinted that blood precursor cells can form other tissues, such as brain cells, if they are first incubated with embryonic stem cells.

The *Nature* papers suggested an alternative explanation. Rather than switching their fate — a phenomenon known as transdifferentiation — the adult cells might actually be fusing with the embryonic cells to become an entirely new type of cell. Fused cells might be too abnormal to be of medical use.

Rudolph Jaenisch, a biologist at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, set the tone of the meeting when he mentioned the two *Nature* papers in his keynote address and



No transdifferentiation: adult bone marrow cells from mice fuse with embryonic cells, resulting in too many copies of each chromosome, according to a team led by the University of Florida's Naohiro Terada.

suggested that all of the evidence for transdifferentiation should now be re-evaluated.

Subsequently, a succession of stem-cell researchers gave talks in which they said they had been unable to confirm the reported versatility of adult cells in their labs. Stanford University biologist Irving Weissman said his group had failed to replicate the transdifferentiation experiments. And Ihor Lemischka, a molecular biologist at Princeton University, told the meeting that stronger evidence for the plasticity of adult stem cells was needed.

Scientists who have reported transdifferentiation said they would look for evidence of fusion, but had no reason to suspect that it accounted for their results. Diane Krause of Yale University, who generated one of the most celebrated adult stem-cell results last

year (see *Cell* **105**, 369–377; 2001), said that her transdifferentiated stem cells contained a normal number of chromosomes, whereas a fused cell should have twice the normal number. Others countered that cells can shed extra chromosomes and might return to the normal number after a few generations.

Adult stem cells may find a use in medicine even if they don't transdifferentiate, says Ronald McKay of the National Institutes of Health in Bethesda, Maryland. He says his colleague Donald Orlic has shown that bone-marrow cells can repair heart-attack damage in rodents, although the mechanism involved is unclear (*Nature* **410**, 701–705; 2001).

Catherine Verfaillie of the University of Minnesota, who is studying stem cells from bone marrow that appear capable of forming any tissue in the body, says she has not yet had time to test whether cell fusion played a part in her findings, but that she would do so.

The fusion argument is likely to come up in the Senate in debates there over a bill introduced by Sam Brownback (Republican, Kansas) that would ban human cloning. The nuclear-transfer procedure used in cloning could also be used to produce genetically compatible embryonic stem cells for individual patients. Brownback has argued that adult stem cells make this unnecessary, as they can be taken directly from the patient. "There is no doubt that fusion will be discussed in the Senate hearings," says McKay.

Brigid Hogan, a cell biologist at Vanderbilt University School of Medicine in Nashville, says that research on adult stem cells will ultimately benefit from the fusion controversy because it will create new standards for proving that transdifferentiation is possible. "Rigorous standards must be set if people are going to talk about using adult stem cells as a clinical solution," she says.

Toxic transfer gets cool response

Virginia Gewin, Washington

Members of Congress and environmentalists are rallying against a proposed transfer of some water-quality research programmes from the US Geological Survey (USGS) to the National Science Foundation (NSF).

The transfer of the Toxic Substances Hydrology Program, proposed by President Bush as part of his budget in February (see *Nature* **415**, 565; 2002), would also involve a \$22-million cut in funding for the programme, which analyses the concentration of biological and chemical pollutants in streams and rivers. A similar cut was proposed by the administration last year, and was rebuffed by Congress.

Representative Ron Kind (Democrat, Wisconsin) has drafted a letter urging the

appropriations subcommittee responsible for the USGS to maintain this year's budget for the water programmes, and to dismiss the transfer. The letter is being circulated to other representatives for signature.

The results of a USGS study released last week — the first to document levels of 82 compounds, including reproductive hormones, in streams across the country — prompted 40 environmental groups to call for the programme to remain at the USGS.

"The USGS is the ideal institution to conduct nationwide surveys such as this one," says Margaret Mellon, head of the food and environment programme at the Union of Concerned Scientists. Critics say that the programme would change beyond recognition if it was transferred to the NSF. ■

Anger at halal law imperils German research

Quirin Schiermeier and Marion Kerstholt

Germany is poised to pass a constitutional amendment that may lead to tighter restrictions on the use of animals in research.

Following a policy reversal by the main opposition party, the constitution is likely to be modified this year to state that animals must be treated as fellow creatures and protected from avoidable pain.

Such an amendment has been backed for several years by the governing Social Democrat and Green parties, but its passage by the required two-thirds majority was blocked by the opposition Christian Democrats.

The Christian Democrat leaders now say they support the amendment. According to party leader Angela Merkel, the change is linked to a controversial judgement by Germany's highest court on halal slaughtering. The court ruled in January that Muslims, like Jews, should be allowed to slaughter animals in this way. In this form of slaughtering, the animal is not stunned before it is killed.

The judgement has been badly received by the German public, and animal-welfare groups such as the Deutsche Tierschutzbund have intensified their campaign to change the constitution. But scientists fear that the proposed bill will interfere with their right to carry out research — which is separately enshrined in the constitution.



Hands off: work on animals such as this specially bred tupaia (tree shrew) may become illegal.

Germany's animal-protection laws are already among the strictest in the world, and until now science organizations have successfully campaigned against additional restrictions (see *Nature* **397**, 461; 1999).

Research organizations criticize the Christian Democrats' shift on the issue. "This will create huge problems for basic and industrial research," says Ivar Aune, a spokesman for the Society for Health and Research, Germany's main lobby group for biomedical research. "The requirement of weighing freedom to research and animal

rights is likely to lead to endless legal arguments and further bureaucratic obstruction to research involving animals."

Animal-protection groups expect the amendment to become a strong legal tool for them. "We will do everything in our power to reduce the number of experiments in Germany," says Wolfgang Apel, president of the Deutsche Tierschutzbund. "The constitutional freedom to research will no longer be enough to legally justify the use of animals in research," he predicts.

Andreas Kreiter, a neuroscientist at the University of Bremen, who uses macaques for his studies of electrical signals in primate brains, has been verbally and physically attacked by activists (see *Nature* **396**, 505; 1998). Kreiter says that the amendment "would further threaten" his work.

But researchers now seem to be resigned to its passage in some form before October's elections. "The halal judgement has sensitized the public to an extent that a constitutional amendment seems unavoidable," says Kuno Kirschfeld, director of the Max Planck Institute for Biological Cybernetics in Tübingen, who deals with animal welfare at the Max Planck Society. "We are very unhappy about this development," he says, "but it looks as if all we can do is campaign for soft wording."

ACTION PRESS/REX FEATURES

Alternative therapies leave US commission divided

Erika Check, Washington

A White House panel has failed to reach agreement on the awkward question of how to deal with alternative medicine.

The majority of the White House Commission on Complementary and Alternative Medicine Policy, set up two years ago by President Bill Clinton, has called for the health department to establish a powerful new office to coordinate research and information on 'alternative' treatments such as homeopathy and acupuncture.

But a minority of members has written a statement of dissent — and is instead calling on the government to do more to highlight scientific scepticism about these therapies.

The Bush administration may choose to ignore the findings, which were received by the health department earlier this month and will be passed on to the White House shortly. But although disparaged by many scientists and doctors, alternative medicine is reported to be used by 42% of Americans (D. M. Eisenberg *et al.* *J. Am. Med. Assoc.* **280**, 1569–1575; 1998).

The therapies' fans include Senator Tom Harkin (Democrat, Iowa), who chairs the

Senate panel responsible for funding public health and biomedical research.

The panel was chaired by James Gordon of Georgetown Medical School, director of the Center for Mind–Body Medicine in Washington. Its majority report calls for the government to increase its support for research into alternative medicine. The report adds that "safe and effective" alternatives should be paid for by federal health programmes, such as Medicaid, and that the government should provide incentives for private companies to research natural therapies that cannot be patented.

But two members — Joseph Fins of the Weill Medical College of Cornell University, and Tieraona Low Dog, medical director of the Tree House Center of Integrative Medicine in Albuquerque, New Mexico — filed a dissenting statement. "We felt we needed to give voice to some of the healthy scepticism that exists with regard to complementary and alternative medicines in many areas of American life," Fins says. He adds that the main report spent too much time advocating its subject and failed to justify its call for more research money.

Fins also says that the National Center for Complementary and Alternative Medicine, a branch of the National Institutes of Health, should fund most research on unorthodox therapies. The majority's report suggests that the government should distribute more research dollars for complementary medicine to several agencies.

Health department spokesman Bill Hall says that it is too early to know what the administration will do with the findings.

AP



Points of view: panel members disagreed over funding for therapies such as acupuncture.

Hints of bone bounties rile fossil hunters

Rex Dalton, Denver

Claims that researchers may have paid tribesmen in Eritrea monetary bounties for fossils and ancient stone tools surfaced at the annual meeting of the Paleoanthropology Society this month.

The allegations concern an extremely valuable fossil site in an isolated area of the East African state, near the village of Buia. It was in this region, known as the Danakil Depression, that an Italian-led team in 1995 found clues of early hominids, including a skull (see E. Abbate *et al. Nature* **393**, 458–460; 1998).

The introduction of bounty payments to a virgin exploration area would have a significantly adverse impact, anthropologists say. When impoverished local residents learn that they can make money by selling artefacts, scavenging can destroy palaeontology sites and make specimen dating more difficult. This has already occurred elsewhere in Africa, in China and in South America.

On 19 March, Lorenzo Rook of the University of Florence told the meeting in Denver, Colorado, that when his team returned to the Buia site in January this year, local tribesmen appeared at the camp offering fossils and stone tools for sale. In previous trips to the region, no tribesman was known to have engaged in this practice, Rook said.

Questioned through translators, the tribesmen said that they had been offered “dollars” by recent explorers, Rook claimed, adding that the tribesmen had also learned a few English-language phrases.

About three months previously, another research team had spent several weeks surveying the Buia region during a contentious research trip (see *Nature* **415**, 463; 2002). That team, led by anatomist Randall Susman of Stony Brook University in New York, and including geologist Roger Smith of the South African Museum in Cape Town, covered territory that Rook’s team had previously worked on.

In January, Eritrean antiquities officials cancelled Susman’s permit to conduct research at the Buia site. In a letter, the officials said that Susman should not have worked in the Italians’ area, and that he failed to notify the government of Smith’s presence on the team and had no authorization to collect fossils and stone tools. The specimens recovered by Susman’s team were deposited at the National Museum of Eritrea.

In an interview, Susman emphatically denied paying any bounties to tribesmen for collecting specimens. Such an allegation is “absolutely, utterly false”, he says. Smith also

denies paying anyone for fossils or tools.

Rook, who eventually obtained an exclusive permit to explore the Buia area after antiquities officials became upset with Susman’s activities, says he has filed a verbal report with Eritrean officials about the bounty issue. The officials could not be reached for comment.

Susman, who has explored African sites for 25 years, says he has given up trying to conduct research in Eritrea. Calling the Eritrean actions “a kangaroo court”, Susman says: “I’m done there. It was a morass — a failure waiting to happen. The Italians can have the joy of me being thrown out.” ■

Babies’ cancer screens ‘not needed’

Alison Abbott, Munich

Screening for cancer seems to be ineffective in saving the lives of babies with nerve tumours, concludes a pilot study of some 1.5 million children in Germany.

The babies were screened for neuroblastomas, the second most common tumour found in children. But the results have raised worries that some toddlers may have been put through gruelling treatment — usually involving surgery and chemotherapy — to remove tumours that might have gone away of their own accord.

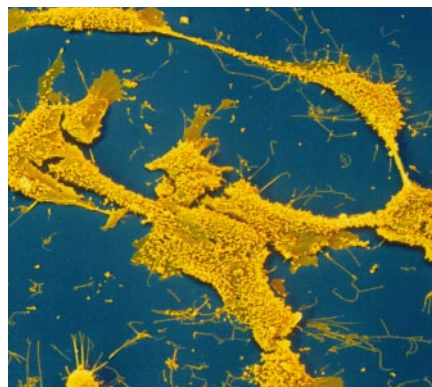
The results of the study were presented as abstracts at the Congress of the German Cancer Society in Berlin earlier this month, and will be published in April in an international clinical journal.

Coming hot on the heels of last October’s assessment of mammography by the Nordic Cochrane Centre in Copenhagen — which raised similar questions about the effectiveness of mass screening for breast cancer — the study is likely to spark fresh debate on cancer-treatment policy.

Neuroblastomas are responsible for the deaths of 15 in every 100,000 children. Chances of survival are improved if the condition is caught early, and the cancer has particular properties that make early diagnosis relatively easy — most tumours produce chemicals called catecholamines, the breakdown products of which can be detected in urine.

The German screening programme was introduced in 1995 with the financial support of Deutsche Krebshilfe, a German cancer charity, and health insurance companies. The parents of about 1.5 million one-year-olds delivered nappies, for urine extraction, to participating clinics in six German states.

The screen identified around 150 cases of



Experts worry that screening children for neuroblastoma cells leads to needless therapy.

neuroblastoma, but missed a further 55 cases, which were later discovered in advanced stages of development. The detection of so many early cancers has not, as yet, reduced the number of deaths from neuroblastoma compared with a control group of unscreened children.

Freimut Schilling, an oncologist at the Olgahospital in Stuttgart, one of the clinics participating in the study, notes that “a considerable number of children may undergo unnecessary treatment on a spontaneously regressing tumour as considerable overdiagnosis was observed”.

Schilling wrote in his abstract that mass screening for neuroblastoma “cannot be recommended as a general health measure”. The authors of the study and officials at the Deutsche Krebshilfe are declining to comment further on the results in advance of their publication — as are scientists participating in the Quebec Neuroblastoma Screening Project, a similar, but smaller, screen in North America that has reportedly obtained similar results. ■



Software firm falls victim to shifting bioinformatics needs

Jonathan Knight, Oakland

DoubleTwist, once one of the most dynamic bioinformatics companies of the dotcom age, died quietly on 7 March at its home in Oakland, California.

At its peak, the software firm beat both public and private DNA sequencers to a count of the genes in the human genome. But its death, after a prolonged decline, was barely noticed outside the world of bioinformatics.

DoubleTwist was born as Pangea Systems in 1993, when Joel Bellsen and Dexter Smith, graduates of Stanford University in the heart of California's Silicon Valley, decided that a market was developing for computer software to manage the explosion of data being produced by geneticists.

Four years later they came up with a business plan that attracted the interest of the valley's top venture-capital firms. By the end of 1997, Pangea had \$7.5 million in the bank and a new chief executive, John Couch.

Couch was a former employee of Apple Computer, and his vision for Pangea Systems, he said at the time, was inspired by Apple's model of user-friendliness. Pangea's mission was to develop tools that would make genome analysis easy for biologists who had better things to do than program computers.

Some Pangea Systems software sold quite well, but the company was up against a stream of free genome-analysis tools flowing out of academic institutions where student programmers had begun doing thesis work

in biology labs. So in December 1999, the company reinvented itself as DoubleTwist.

DoubleTwist.com was Pangea's new web portal, which allowed anyone with a browser to probe the public genome data. According to old Pangea press releases, it was originally intended to be free, to showcase the company's genome-crunching ability. But when it became DoubleTwist, the company moved the portal to the core of its business, charging up to \$10,000 a year for subscriptions.

At the same time, the company was secretly collaborating with Sun Microsystems, a computer company based in Palo Alto, California, to scour the public human-genome database for genes. On 8 May 2000, less than two months before the announcement at the White House of plans to publish the results of the public and private sequencing endeavours, DoubleTwist made headlines by announcing that it had discovered 105,000 genes in the genome.

Joseph Alper, who used to work for DoubleTwist, remembers it as an exuberant time. The company, for example, took its 200 employees on an outing to a horse race meeting. "It was typical of the dotcom attitude before the bubble burst," he says.

In September 2000, the company registered for a public offering on the stock market, but withdrew the offering a few months later, having missed the market's high point. Despite a few software deals, DoubleTwist



Final twist: the shutters are up at a company that sought to gain from repackaging genomic data.

never obtained the income it needed to survive. By January, its staff had been cut to 60.

"Most genome informatics is done in-house, in both industry and academia," says Sean Eddy, who writes gene-hunting programs at Washington University in St Louis, Missouri. Incyte Genomics, of Palo Alto, and Celera, the most famous of the genome information companies, have both seen the writing on the wall, and have sought to recast themselves as drug-discovery firms.

These changes have shaken investors, but the bioinformatics business is not dead, says Stephen Lincoln of InforMax, a biotech software firm based in Bethesda, Maryland. "Clearly we are going through a shakeout phase right now," he says. ■

Baiting plan to remove fox threat to Tasmanian wildlife

Carina Dennis, Sydney

The deliberate introduction of foxes into Tasmania, the island off the southern coast of the Australian mainland, is threatening to devastate the island's unique ecology, wildlife biologists say.

At a conference in the Tasmanian city of Launceston earlier this month, fox and pest-control experts called for steps to eradicate the foxes before the breeding season begins in July. Unless action is taken before then, they say, the island's fox problem could get out of control.

Tasmania is the last refuge for a long list of species that have been wiped out in mainland Australia. Many of these are vulnerable to predation by foxes.

"The information that the authorities have received leaves no doubt that foxes were deliberately brought into Tasmania," says Nick Mooney, a member of the state

government's fox task force, which was set up last year when the problem was first identified. It is believed that the foxes were deliberately imported, perhaps by individuals who wanted new game to hunt or who were angered by the recent introduction of stricter gun-control laws. A police investigation is under way.

There are estimated to be at least a dozen foxes in the state, says Tony Peacock, chief executive of Australia's Pest Animal Control Cooperative Research Centre. "Unless action is taken to eradicate the foxes before the breeding season starts, the battle to keep Tasmania fox-free may be lost," says Peacock, who says that an estimated 77 native species could be in jeopardy as a result.

Farmers, biologists and conservationists at the Launceston meeting agreed that the foxes should be eliminated by baiting them with sodium monofluoroacetate — a poison

better known to farmers as '1080' — in conjunction with shooting and trapping. Use of the bait is highly controversial in Tasmania, where conservationists object to farmers and foresters using it to control populations of native animals, such as opossums and wallabies, which graze on crops and tree seedlings. Biologists and wildlife managers are trying to work out how to present the bait in a form that will be accessible to foxes, but not to other native carnivores, which include marsupial species such as quolls and the Tasmanian devil.

Closer to breeding time, Roger Short, an expert in reproductive biology at the University of Melbourne, hopes to test the use of sterilized vixens carrying oestrogen implants — which keep them in constant oestrus or 'heat' — to lure male foxes, which will then be killed. ■

Genomics projects hint at Europe's future priorities

Brussels The European Commission this month unveiled three genomics research projects, giving researchers a good indication of the direction European funding for their field will take in coming years.

Like many of the initiatives under the commission's 2.2-billion-euro (US\$1.9-billion) human genomics programme, the projects will involve large networks of scientists from public institutions and private companies across the continent.

One consortium will conduct structural studies of 600 human proteins that are thought to be medically important. Another will do epidemiological research into twins as a way of studying the genes responsible for common diseases. The third project centres on mouse genomics. Peter Gruss, director of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, says that this is the first time the commission has funded a systematic study of the mouse genome. Funding for the three projects will total 40 million euros.

Attempt to uproot refugees from lab ends in violence

Moscow A laboratory and library at the Botanical Institute in Tbilisi, Georgia, have been destroyed after police tried to evict refugees occupying a building on the institute's grounds.

The refugees took over the building in the past few weeks after fleeing a conflict in the

republic of Abkhazia, a mountainous region with a population of 120,000, which declared independence from Georgia in 1993. News agencies reported that the refugees had rebuffed repeated requests by the institute's director, Georgy Nakhutsrishvili, for them to leave. Local police and special forces from the Ministry of Justice attempted to remove the refugees on 20 March, prompting a riot in which laboratory property was smashed and rare library books were burned.

Shields to make light work of city pollution

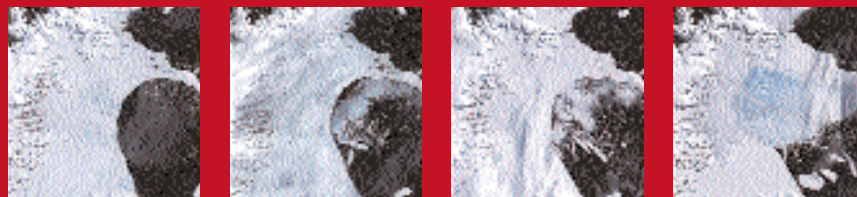
Prague The Czech Republic has become the first country in the world to legislate against light pollution, a serious problem for astronomers working near towns and cities.

Czech city administrations have only two months to adapt their street lights, the main source of light pollution, to meet the new regulations. The task is a mammoth one. "In Prague the whole street-lighting system needs to be equipped with fully shielded light fixtures," says Jenik Hollan, an astronomer at the Nicholas Copernicus Observatory and Planetarium in Brno, who was involved in preparing the legislation.

Astronomy-friendly lights are rare elsewhere. Tucson, Arizona is one of only a few cities to have converted its street lights so that they do not emit light upwards.

Bushmeat could be bridge to humans for HIV's cousins

Paris More than one-fifth of the monkey meat sold in the West African nation of Cameroon is infected with HIV's ancestor, the simian



Satellite snaps a shortened shelf life

Boulder The dramatic disintegration of an Antarctic ice shelf, believed to be about 12,000 years old, has been captured by one of NASA's Earth-observation satellites.

Researchers working with data from the Terra satellite were already aware that the Larsen B ice shelf was gradually breaking up. But this sequence of pictures, taken over the course of five weeks, reveals a rapid acceleration in the process, with over 3,000 square kilometres of the 200-metre thick shelf breaking free since January.

Climate change is the most likely cause. Polar researchers say that Antarctica has warmed by over 2 °C in the past 50 years, creating countless pools of meltwater on the ice sheets. These seep into cracks in the sheet, eventually leading to the breakup of the shelf.

Researchers are now trying to learn more about the breakup by studying other ice shelves. They hope that climate models will clarify the processes behind the rise in Antarctic temperatures. "Global climate modellers now have to determine what forcing on the climate resulted in this warming," says Theodore Scambos, a geologist at the University of Colorado in Boulder.

J. DESCLOITERS/MODIS/NASA

immunodeficiency virus, scientists from the Research Institute for Development in Montpellier, France, have found. The findings highlight the risk of new HIV-like viruses infecting humans who eat bushmeat, the researchers say.

The team screened blood samples from 16 species of monkey and ape, revealing 21 types of the simian virus, four of them new to science (M. Peeters *et al. Emerg. Infect. Dis.* 8, in the press). The range of strains found is worrying, as the chances of infection are thought to increase if a person is exposed to a greater number of strains.

The Montpellier researchers are now sequencing the genomes of the strains they have collected, and aim to develop tests for the viruses. They then hope to screen people who prepare or eat bushmeat to see which, if any, strains they are carrying.

► www.cdc.gov/ncidod/eid

Women find it's still a man's world at MIT

Boston A large increase in the number of female members of faculty at the Massachusetts Institute of Technology (MIT) has failed to prevent women at the institute from feeling marginalized, a new report has found.

The institute first began investigating the issue of gender bias in 1994, since when the number of women in teaching positions has leapt by 40%. But its latest study, a follow-up to a 1999 report that revealed a systematic pattern of gender discrimination within some departments (see *Nature* 398, 361; 1999), admits that a bias continues to exist.

Relatively few women hold positions of power at the institute and, although salaries for female faculty members have increased in some departments, many female staff continue to feel that they wield less influence than their male counterparts, the authors say. There is a pervasive feeling, writes provost Robert Brown in the report's introduction, "that MIT is a 'man's world'. This must change".

► web.mit.edu/faculty/reports

The heady world of soccer nets a clean bill of health

Washington Worried parents can relax—playing soccer might not cause the cognitive impairments suggested by the results of recent studies.

The American Academy of Pediatrics said in March 2000 that the potential risks of heading a soccer ball, or of clashing heads with other players, were not known. As a result, the academy called on youth soccer organizers to minimize heading.

But a new study from the University of North Carolina at Chapel Hill, in which



On the ball: there are no penalties for footballers who tackle their job head-on.

researchers studied the cognitive abilities of soccer players and other athletes at the university, found no evidence of reduced performance on standard cognitive tests (K. M. Guskiewicz, S. W. Marshall, S. P. Broglio, R. C. Cantu and D. T. Kirkendall *Am. J. Sports Med.* 30, 157–162; 2002).

But the researchers did discover that soccer players suffer more concussions than other athletes, and critics warn that the university's entrance requirements may have weeded out soccer players who have already suffered brain damage.

Oil drilling sends cod into sexual decline

Bergen The North Sea's dwindling cod stocks could be facing a new threat. Researchers in Norway have found that chemicals released by oil drilling can disrupt the ability of the fish to reproduce.

In laboratory experiments, male cod took on female characteristics and female fish spawned later and produced smaller eggs when exposed to alkylated phenols, which occur naturally in underground oil reservoirs.

Scientists from the Norwegian Institute of Marine Research announced their findings last week at a conference on the North Sea, held in Bergen in Norway, and are now carrying out field trials around drilling platforms.

A separate report from the same conference warned that the common skate is on the verge of extinction in the North Sea because of overfishing. A ban on sea-bed trawling in certain regions is the only way to protect bottom-dwelling species such as skate and rays, the report's authors say.



Can you believe what you read?

Scientists' financial interests can bias the papers and review articles that they write, studies suggest. But what can editors do to police the issue? Frank van Kolschooten examines journals' policies on conflicts of interest.

Money talks, so they say. But is it whispering covertly through the pages of leading science journals? Too often, the answer is yes, claims a letter sent last month to some 200 journals by the Center for Science in the Public Interest (CSPI), a non-profit organization based in Washington.

The letter urged the journals to strengthen their policies on the disclosure of conflicts of interest — and given that its signatories include former editors of *The New England Journal of Medicine* (NEJM) and *The Journal of the American Medical Association* (JAMA), it cannot simply be dismissed as the blusterings of a fringe element. “Whether the issue is clinical research, cancer clusters or global warming, corporate interests can hide

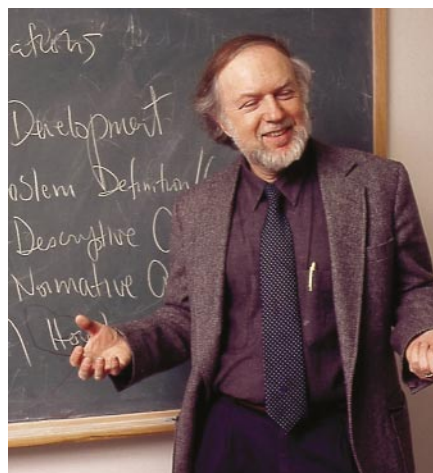
behind the credibility of peer-reviewed journals,” argues Virginia Sharpe, who heads the CSPI’s Integrity in Science project.

As the links between commerce and academia deepen, how to deal with the conflicts of interest that inevitably arise has become an increasingly important issue. Next week sees two meetings devoted to the subject: at Emory University in Atlanta, Georgia, researchers, entrepreneurs and others will meet to debate the ‘Commercialization of the Academy’; meanwhile, Warsaw in Poland will host an international conference on conflicts of interest in science and medicine.

The Warsaw meeting will be addressed by the current editors-in-chief of *JAMA* and *NEJM*. As research agencies, academic institutions and scientific societies all debate

the issue of conflicting interests, journal editors are finding themselves on the front line — the scientific literature is, after all, the main forum for the communication of research results. Some of these editors are thinking hard about strategies to minimize the potential for publications to be biased by commercial pressures, and to draw their readers’ attention to any conflicts of interest that may exist.

Lisa Bero, a pharmacologist at the University of California, San Francisco, who studies how science influences clinical practice, has no doubt that commercial interests are biasing the scientific literature. “Studies that are sponsored by a single company are biased compared with studies with multiple or other sponsors,” argues Bero, who signed the CSPI



Lisa Bero (left) and Sheldon Krimsky worry that vested interests are distorting published science.

letter. "When research is funded by one company that has an interest in the outcome it is much more likely to have a favourable outcome for the sponsor's product."

In 1986, Richard Davidson of the University of Florida College of Medicine in Gainesville reviewed 107 published clinical trials and found that those sponsored by drug firms were more likely to report favourably on the treatment being tested¹. Since then, several studies have provided support for Davidson's conclusion². One famous example reviewed 70 articles commenting on the safety of calcium-channel antagonists, a class of drugs used to treat cardiovascular disease. Among the authors of original research papers, reviews and letters to the editor that were supportive of the drugs' use, 96% had financial relationships with the drugs' manufacturers; for publications deemed neutral or critical the figure was only 60% and 37%, respectively³.

Clinical precision

The factors that underlie such biases remain unclear. "It could be because negative or unfavourable studies aren't published; it could be that companies are not dumb enough to fund a study that is not going to work out; it could be that they are conducted or designed in a poor way," Bero speculates.

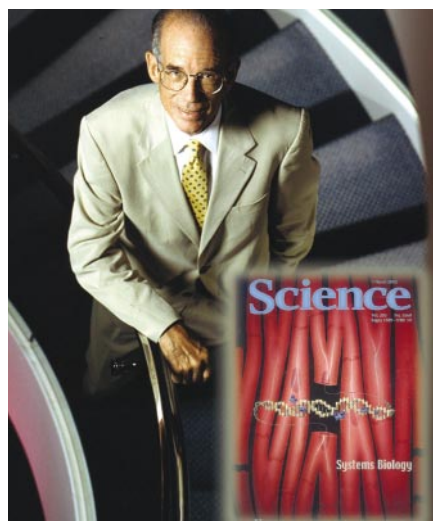
Concerns about commercial conflicts have been most acute in clinical medicine — where human lives may be at stake. Medical journals have led the way in introducing editorial policies to deal with conflicting interests, and they continue to blaze a trail (see 'Who controls the data?', overleaf). At the heart of each policy lies the concept of disclosure: if everyone is made aware of authors' financial interests, goes the argument, the potential for bias can be borne in mind by the reader. But how full that disclosure should be, and how to encourage authors to comply with a stated policy, remain matters for debate.

Leading multidisciplinary journals such as *Nature* and *Science* have also adopted con-

flict-of-interest policies — as have journals in fields such as nutrition, where commercial conflicts have become a serious concern. But in many scientific disciplines the issue is still not on editors' radar screens, despite the best efforts of the CSPI and its supporters. "This is a potential problem in any science in which there is a commercial interest," says Sheldon Krimsky of Tufts University in Medford, Massachusetts, who has studied the conflict-of-interest policies of leading journals and will speak at next week's meeting in Atlanta.

The CSPI's letter was sent to journals in fields including climate research, environmental science and chemistry. Bruce Coull of the University of South Carolina in Columbia was moved to sign the letter by his concerns about commercial conflicts in his own discipline of marine ecology. "Many scientists in our field are sponsored by companies or work as consultants," he says. "I would like to know that when I read their publications in ecotoxicological journals, but most of these journals don't have disclosure policies."

Orrin Pilkey, a coastal geologist at Duke University in Durham, North Carolina, and



another signatory of the CSPI letter, feels the same way. "I see more and more examples of studies with a huge amount of optimism, written by scientists who also work as consultants for companies," he says.

But it seems that the CSPI has some way to go in convincing journal editors to address the issue. "Life is pretty complicated already and this would be another layer of paperwork," says Charles Finkl of Florida Atlantic University in Boca Raton, and editor-in-chief of the *Journal of Coastal Research*. Other editors fear that rigid policies could make researchers send their papers elsewhere. "Every time you put another barrier in the way, people can go to another journal," observes Martin Blume, editor-in-chief of the American Physical Society, which publishes *Physical Reviews*, *Physical Review Letters* and *Reviews of Modern Physics*.

Blurred vision

Even among editors who have embraced the general principle of disclosure, there is considerable divergence over what this should mean (see table, overleaf). One controversial issue, for instance, is whether authors' conflict-of-interest statements should be given to the referees asked to review a manuscript. Donald Kennedy, editor-in-chief of *Science*, explains why his journal does not do this: "That's not part of the referees' job. They are supposed to review the quality of the science; they are not ethicists." *Nature* has the same policy of not disclosing conflict-of-interest statements to reviewers.

But some journals do send the disclosure statements to referees — for good reason, argues Krimsky. "Personal financial interest is a relevant factor in raising the scepticism of peer reviewers," he claims. "There are too many articles that are not sufficiently rigorously reviewed, so the flaws in them are only disclosed after publication."

Another issue that divides the journals is how to deal with the financial interests of the referees themselves. Jeffrey Drazen, editor-in-chief of *NEJM*, says that his journal pays close attention to this potential source of bias. "We ask



In agreement: Donald Kennedy (left) and Philip Campbell back disclosure to readers but not referees.

Who controls the data?

Last September, 11 major medical journals introduced a tough new publication policy. In future, they demanded, authors must sign a statement verifying that they have been involved in the design of the study, and in analysing and interpreting the data. They also must declare that they have seen the raw data and that the sponsor of the research did not have control over whether the results could be published.

"This declaration concerns all kinds of sponsors," says Catherine DeAngelis, editor of *The Journal of the American Medical Association* (JAMA). "But it was the pharmaceutical companies that forced us to do this. We know several examples of studies where the company and not the sponsored researcher controlled the data."

In several notorious examples, companies have suppressed the publication of clinical trials that reflected poorly on their products. In the 1990s, for instance, a team led by Nancy Olivieri of the Hospital for Sick Children in Toronto, Canada, concluded that deferiprone, used to treat thalassaemia, did not adequately control the build-up of iron in patients' livers. Thanks to a clause in Olivieri's contract with the drug's manufacturer, Apotex, she had to wait three years before publishing her results⁸.

In another case, results suggesting that Synthroid, the leading drug used to treat people with underactive thyroid glands, was no more effective than cheaper alternatives remained unpublished for seven years. During this time, the British company Boots Pharmaceuticals tried to enforce a clause in its contract with Betty Dong of the University of California, San Francisco, that gave the company a veto on publication. After a titanic tussle, Dong eventually published her results in *JAMA*¹⁰.

How many more such papers have never seen the light of day, and how many favourable papers are really the work of drug companies, rather than the scientists named as authors, remains unclear. But Jeffrey Drazen, editor-in-chief of *The New England Journal of Medicine*, says that his experience of researchers who fail to answer questions about their own manuscripts convinces him that the latter does occur. "They don't call you back themselves, but you get the sponsor on the phone, who doesn't want to tell you everything because of his competitive position," says Drazen. "Usually we are forced to reject such a manuscript."

Richard Horton, editor of *The Lancet*, says that the reaction from industry to the new rules has been interesting. "First there was massive condemnation," he says. "But a few months later the reaction was that they were going to comply with all our recommendations because they want to be seen doing the right thing."

In the case of clinical trials, argues Horton, taking the 'right' course of action is not simply an abstract moral issue. "There are too many examples of drugs which had been licensed that had to be withdrawn because the supporting data were inadequate or because the company put too much positive spin on weak data," he says.



Jeffrey Drazen: referees may introduce bias.

negative and positive remarks."

Nature asks referees to disqualify themselves if they think judging a certain manuscript would create a conflict of interest. "But we do realize that it can happen that a referee judges a manuscript he or she should have pushed aside," says *Nature's* editor, Philip Campbell. "We often use three referees, which helps to avoid such problems."

Opinion toll

Other journals that demand disclosure of conflicts by authors have no formal policy for the interests of referees — the *American Heart Journal* is one example, although some of its referees have, on occasion, declined to review a manuscript, citing a conflict of interest.

The authorship of opinion and review articles has also emerged as a contentious issue. Although most journals adopt the same rules as for original research papers, *NEJM* has, since 1990, required authors of such pieces to have had no commercial ties for at least two years with companies whose products they are writing about. But the policy is becoming increasingly difficult to enforce, admits Drazen, and is now under review. "Since the introduction of that rule, the entanglement of scientists and companies has grown," he says. "Therefore we have to exclude a lot of scientists from writing comments. We have collected data to see if that's still the right policy for the journal."



Catherine DeAngelis: no stocks, no conflict.

she took up her post in 2000. "I am very careful," she says. "When I received a call that my uncle had left me stocks in the pharmaceutical company Johnson & Johnson, I said immediately that they should be given to my sisters. I don't want stocks."

Drazen, an asthma researcher at Harvard

them to provide information that throws a light on possible commercial and intellectual prejudice. When a referee gives negative advice about a study we liked very much ourselves, we make some calls," he says. "We have seven editors spending all their time reading reviews and judging the relevance of

There are journals where the financial ties of the editor will determine what gets published.



Mildred Cho

University, had to sever his financial ties to 20 drug companies before he took over the helm at *NEJM* in 2000. "The proceeds from stock investments I gave to charity," he says. Drazen also agreed not to deal for two years with manuscripts involving products from companies in which he previously had a financial interest.

Nature requires editorial staff to declare to their managers any interests that might be perceived to influence their editorial judgement. Managers then decide how any potential conflicts should be dealt with.

But move beyond front-line journals, and things become rather murky. "There are journals where the financial ties of the editor will determine what gets published," claims Mildred Cho, a bioethicist at Stanford University in California. She is particularly concerned about journals publishing papers about medical devices. "After publication, those papers are used as marketing tools by the companies that produce the devices," says Cho. "But we don't know the prevalence of editorial staff having commercial ties."

Hidden agendas

Even if a journal has a clear conflict-of-interest policy, it is of limited use if it is widely ignored. Unfortunately, it seems that this is often the case. Last year, Krimsky published a study of articles appearing in 1997 in 1,396 high-impact journals⁴. Only 15.8% of them had an explicit conflict-of-interest policy, of which almost 90% were medical journals. Among the journals with a stated policy, only 0.5% of papers included a disclosure of conflicting interests, and 65.7% of these journals published zero disclosures. In an earlier study, Krimsky unearthed evidence of lead authors with relevant commercial interests in 34% of a sample of 789 papers⁵, so he does not believe that so few authors had conflicts to declare. "Poor compliance is the more likely explanation," says Krimsky.

Nature introduced its conflict-of-interest policy in October 2001. Again, disclosure rates are relatively low: of the first 110 papers accepted under the policy in 2002, only five included a declaration of a financial interest.

Staff at *NEJM* know from bitter experience all about the subtle difficulties of policing a



Richard Smith: sees a rise in declarations.

conflict-of-interest policy. Following revelations in the *Los Angeles Times* that some authors of review articles in *NEJM* had commercial interests in the treatments they were writing about, the journal conducted an internal review. In February 2000, *NEJM* revealed that, since January 1997, 19 of the 40 drug-therapy review articles it had published were written by scientists with industrial links that should have disqualified them under the spirit of the journal's strict policy⁶. The authors had slipped through a loophole that exempted financial support given to their institution, rather than to them as individuals.

Richard Smith, editor of the *British Medical Journal (BMJ)*, agrees that it is hard for journals to enforce their policies. "A lot of researchers still think they are immune to the influences of their sponsor and don't realize

that bias can slip into their research very subtly," he says. "Some see it as an infringement on their freedom. But I must say the numbers of disclosure statements we get are increasing. Maybe the culture has started to change."

Cultural revolution

A recent study by Smith reinforces that view. Counting the disclosures in editorials, original research papers and letters to the editor in five major medical journals — the *BMJ*, *JAMA*, *NEJM*, *The Lancet* and *Annals of Internal Medicine* — in 1989, 1994, 1996 and 1999, Smith found an increase, from two declarations in 1989, eight in 1994 and four in 1996, to 38 in 1999 (ref. 7). But still, the vast majority of the 791 articles published by the journals in 1999 contained no disclosures.

Some researchers who have become embroiled in rows over conflicts of interest agree that disclosure is to everyone's benefit. In October 1997, toxicologist Stephen Safe of Texas A&M University in College Station wrote an editorial for *NEJM* in which he argued that environmental oestrogens such

as polychlorinated biphenyls do not cause breast cancer, and attacked public "chemophobia" fed by "paparazzi science"⁸. When it emerged that Safe had previously received funding of \$150,000 from the Chemical Manufacturers Association, the editorial became mired in controversy.

Safe defends his integrity: "My views on endocrine disruptors have been fairly consistent over the years and based on the scientific data; needless to say, I was upset over the commotion." But in retrospect, he now agrees with the idea of full disclosure of current and prior funding sources.

As to whether well-enforced disclosure policies would reduce commercial biases in the scientific literature, no one can say for sure. "Transparency about conflicts of interests is a bare minimum," argues Bero. "People shouldn't have the idea that everything is OK just because financial conflicts of interest are disclosed."

Krimsky notes that disclosure means different things for different journals — some merely require authors to tick a box indicating whether they have financial interests, whereas others demand a full breakdown of what those interests are. He argues that investigating the effectiveness of different types of conflict-of-interest policy is an important avenue for future research — but says it is difficult to get funding for studies in this area.

Many important questions remain unanswered, agrees Bero. "Do disclosure policies discourage investigators from submitting to journals? How do readers use conflict-of-interest information — do they take it into account when reading an article? Are disclosures that are published in journals accurate?"

Given the increasing number of papers getting bogged down in accusations of commercial bias, perhaps it is time to start searching for some answers. ■

Frank van Kolfschooten is a medical writer in Amsterdam.

1. Davidson, R. A. *J. Gen. Intern. Med.* **1**, 155–158 (1986).
2. Bodenheimer, T. *N. Engl. J. Med.* **342**, 1539–1544 (2000).
3. Stelfox, H. T., Chua, G., O'Rourke, K. & Detsky, A. S. *N. Engl. J. Med.* **338**, 101–106 (1998).
4. Krimsky, S. & Rothenberg, L. *S. Sci. Eng. Ethics* **7**, 205–218 (2001).
5. Krimsky, S., Rothenberg, L. S., Stott, P. & Kyle, G. *S. Sci. Eng. Ethics* **2**, 395–410 (1996).
6. Angell, M., Utiger, R. D. & Wood, A. J. *N. Engl. J. Med.* **342**, 586–587 (2000).
7. Hussain, A. & Smith, R. *Br. Med. J.* **323**, 263–264 (2001).
8. Safe, S. H. *N. Engl. J. Med.* **337**, 1303–1304 (1997).
9. Olivieri, N. F. et al. *N. Engl. J. Med.* **339**, 417–423 (1998).
10. Dong, B. J. et al. *J. Am. Med. Assoc.* **277**, 1205–1213 (1997).

Integrity in Science project

♦ www.cspinet.org/integrity

Commercialization of the Academy conference

♦ www.emory.edu/PROVOST/SamNunnForum

International Conference on Conflict of Interest and its Significance in Science and Medicine

♦ surfer.iitd.pan.wroc.pl/events/

ConferenceApril2002.html

Nature's policy

♦ www.nature.com/nature/submit/competing/index.html

Table: **Sample policies from leading journals**

	Authors asked to disclose financial interests?	Disclosure statements shown to referees?	Disclosure statements published?	Policy for referees	Policy for editorials/ review articles	Rules for editors
<i>British Medical Journal</i>	Yes	Yes, if received before paper goes to review	Yes	Asked to declare conflicting interests; if editors perceive a conflict, may seek alternative referee	Same as for research papers	Editors asked to declare interests; expected not to make decisions on papers in which they have an interest
<i>Journal of the American Medical Association</i>	Yes	No	Yes	Asked to declare conflicting interests, and to disqualify themselves if these might cause bias	Same as for research papers	Editors and editorial board members must sign financial disclosure statement; expected not to make decisions on papers in which they have an interest
<i>Lancet</i>	Yes	Yes, if statement included in manuscript, as instructions to authors now request	Yes if editors perceive that a conflict may exist	Asked to declare conflicting interests; editors may exclude referees if they perceive a conflict	Same as for research papers, but editors may avoid commissioning from authors with declared interests	Staff invited to declare financial interests to editor-in-chief; editors excluded from papers in which they have an interest
<i>Nature</i>	Yes	No	Yes	Asked to disqualify themselves if they perceive a conflict	No disclosure policy at present; under consideration	Staff must declare financial interests to their managers, who decide how potential conflicts should be dealt with
<i>New England Journal of Medicine</i>	Yes	Sponsorship and author affiliations only	Yes	Asked to declare conflicting interests; this information taken into account by editors	Authors with conflicting interests are excluded	Editors must have no conflicting financial interests
<i>Proceedings of the National Academy of Sciences</i>	Yes	Yes	Yes	Asked to disqualify themselves if a conflict exists	Same as for research papers	Editorial board members asked to disqualify themselves from handling papers if a conflict exists
<i>Science</i>	Yes	No	No, except in rare cases at editors' discretion	Asked to declare conflicting interests; editors may exclude referees if they perceive a conflict	Same as for research papers	Editors expected to follow same guidelines as referees

The sight of two brains talking

By simultaneously scanning the brains of contestants playing a simple game, researchers aim to study how social interactions affect brain activity. Steve Nadis meets the scientists who think two heads are better than one.

At the crux of a difficult negotiation, have you ever wanted to peer inside your opponent's head to see what he or she is thinking? If so, you are not alone. Around two years ago, during a conversation in an ice cream shop in Houston, two neuroscientists speculated about the activity in the brains of contestants playing a simple game. Might the neural activity in the players' brains be linked, they wondered?

This summer, those researchers will take a step towards answering that question. Two subjects — one in Atlanta, the other in New Jersey — will play a strategic game over the Internet, while having their brains scanned. The team calls the technique 'hyperscanning'. "Virtually everything we do in real life is motivated by interactions with others," says Gregory Berns of Emory University in Atlanta, Georgia, one of the project's masterminds. "It makes sense to probe the biological circuits underlying these interactions."

Berns is interested in the brain activity of subjects playing the Prisoner's Dilemma, a two-person game that is often used as a simple model of real-world negotiations. He wants to see if there is a particular pattern of brain activation that precedes one player's decision to betray the other.

Head start

Berns has already used functional magnetic resonance imaging (fMRI), which tracks neural activity by monitoring blood flow in the brain, to study individual players. But monitoring them one at a time was of limited use. "You could study this with only one person in the scanner, but in this game you never know who will initiate such a move," he says. To study properly how patterns change during the contest, Berns says that simultaneous scanning is needed.

Together with Read Montague, a neuroscientist at Baylor College of Medicine in Houston, Berns carried out a proof-of-principle experiment in November 2000. Two subjects played a simple, computer-based guessing game, while fMRI scans were taken. Montague says that this "toy experiment" was not meant to produce hard results, but it did reveal evidence of synchronized activity in the one area of the players' brains — the inner prefrontal cortex.

The team is now preparing an experiment in which contestants at different locations will compete over the Internet. "Scanners are

expensive and no one owns a lot of them," explains Sam McClure, a graduate student in Montague's lab. "You have to rely on the Internet." One subject will be at Emory, the other at Princeton University in New Jersey, where cognitive scientist Jonathan Cohen is collaborating with the experiment.

Mind games

Months of effort have gone into developing the equipment needed to synchronize the scans and the game, in which the players will work together to guide a hunter to his prey. A hyperscanning control room is being built at Baylor's neuroimaging laboratory. "Initially, we'll run the experiments from Houston — just like a spaceshot," quips Montague.

He and Berns are confident that their technique will attract other users. Lane Strathearn, a developmental paediatrician at Baylor, has studied neural activity in mothers looking at pictures of their children and other infants, and now hopes to scan mothers and children simultaneously to see whether their brains show synchronous activity.

Economist Colin Camerer of the California Institute of Technology in Pasadena is also excited about the new technology. He says that understanding the brain activity underlying the Prisoner's Dilemma could shed light on the tactics used by stock-market traders.

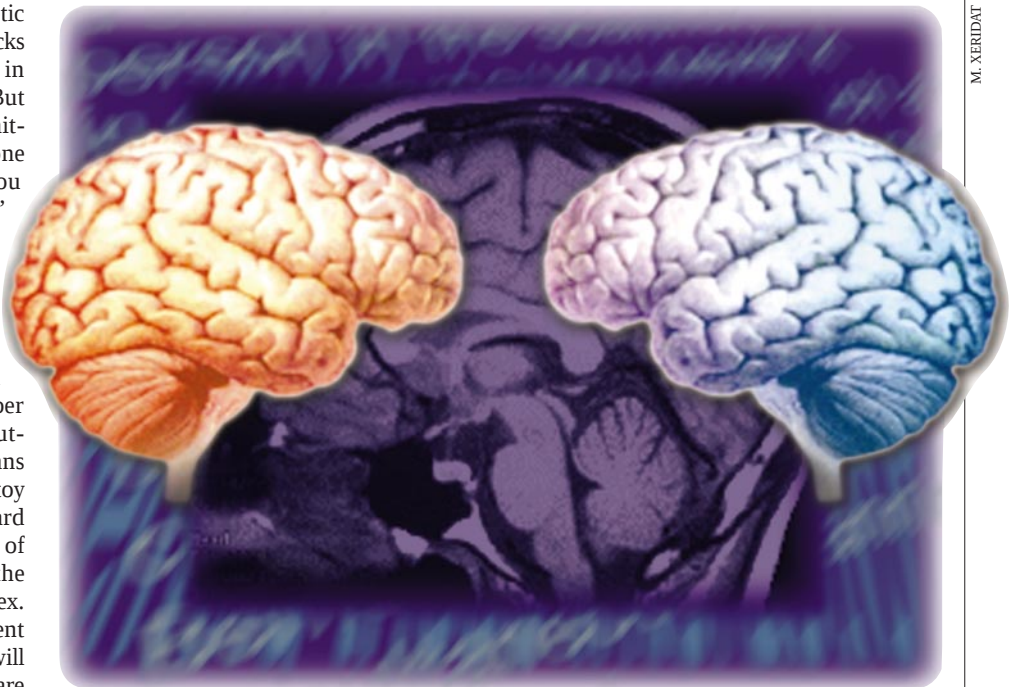


Man with two brains: Read Montague aims to bring linked neural activity out of the shadows.

Other neuroscientists are less enthusiastic. Stanford researcher David Heeger uses fMRI to study visual perception, and is currently investigating strategy and decision-making. He says he will hook two scanners together if someone can convince him that the data need to be collected simultaneously. So far, he says, no one has made a compelling case.

But hyperscanning's advocates are confident that applications will become apparent. "We're hoping that all kinds of people will suggest worthwhile things to do," says McClure. "The same thing happened with conventional brain imaging. In the early days, no one imagined all the things it would be used for." For McClure and his colleagues, it's a case of 'if we build it, they will come'.

Steve Nadis is a freelance writer in Boston.



The rewards of science extend far beyond publication

Good research and wise mentorship should be valued more highly than a name added to a paper.

Sir—Peter A. Lawrence, in his Commentary article “Rank injustice” (*Nature* 415, 835–836; 2002), describes ways in which senior scientists routinely abuse and exploit their juniors. My own experience is very different.

When I was a graduate student and postdoc, my mentors spent countless hours discussing science, experiments and data with me. When they gave seminars, my mentors advertised my work and acknowledged my contributions with every slide. As a result of all their guidance, I have been fortunate enough to have obtained my own lab, to have thrived in it, and to now have the great honour and pleasure of training young scientists myself. In short, good mentors spend time training their students, credit them fairly for their work, and guide them to independence; bad mentors do not.

Young scientists would be well advised to seek out good mentors. Our training system may not be perfect, but when I look around my department and university, I see that quality mentorship is unfailingly taken as seriously as is doing good science. For Lawrence to claim that mentors are routinely abusing young scientists is as irresponsible as it is cynical. More than obtaining fair “allocation of credit”, the reward of doing good science includes learning about nature, helping other people and solving mysteries.

How do I rate other scientists? I don’t count their *Nature* papers but rather how many of their students do well in their own labs. Now that’s success!

Ben A. Barres

Department of Neurobiology,
Stanford University School of Medicine,
Stanford, California 94305-5125, USA

The perils of putting career before all else

Sir—I congratulate you for the stimulating Commentary article about bad mentorship by Peter A. Lawrence (*Nature* 415, 835–836; 2002). While I am a bit pessimistic about addressing this issue solely from the perspective of the mentors, I can suggest a remedy that can help improve the situation almost immediately: students and fellows must evaluate the mentorship potential of labs they are looking at with an eye towards more than the relative ‘fame’ of that lab.

It is appropriate to consider the manner in which trainees are mentored in a variety of ways before joining a lab. If this decision is taken solely on career considerations and not scholarship, will it be a surprise to find that the lab, and the mentor, are dominated by blatant careerism?

Bruce Spiegelman

Dana-Farber Cancer Institute,
Harvard Medical School, 44 Binney Street,
Boston, Massachusetts 02115, USA

A quizzical view pays clear dividends

Sir—It is always a pleasure to read one of Peter Lawrence’s polemics¹. On this occasion he is, of course, correct. Honorary

authorship is a bad thing, especially for younger scientists—that is, younger than Lawrence. I am now making a habit of publicly quizzing authors whose position is clearly an honorary one, about some abstruse technical detail of ‘their’ papers. This is a course of action I recommend.

I cannot, however, avoid admitting that I too have, on occasion, been an author of a paper on which my position as such was wholly unwarranted. One of the most embarrassing examples is a paper—albeit of little note—in which I was invited to be a co-author by one so dear to me that I could not refuse². I cite it here, because I do not believe many others have.

Michael Ashburner

Department of Genetics, University of Cambridge,
Downing Street, Cambridge CB2 3EH, UK

1. Lawrence, P. A. *Nature* 415, 835–836 (2002).

2. Lawrence, P. A., Ashburner, M. & Johnston, P. *Genetics* 134, 1145–1148 (1993).

Film industry shows how to give credit where due

Sir—Few would dispute that the authorship system we have today is open to abuse and probably is frequently abused, as described by Peter Lawrence in his Commentary article (*Nature* 415, 835–836; 2002). Nevertheless, a long, successful stint in research laboratories is

required before anyone gets to be a principal investigator (PI), during which time the researcher has built up a corpus of knowledge and experience that benefits the younger scientists who come to work under his or her supervision.

Even if the PI is not the “discoverer” in a given instance, it is the PI who has created the circumstances in which the discovery was made. The PI deserves credit for qualifying herself or himself to head a laboratory, selecting the problem, getting the funding, recruiting younger colleagues and contributing to the work in various direct and indirect ways.

Rather than taking tennis rankings as a model to avoid, why not take a leaf out of the film industry’s book? After every film, the credits not only name contributors but also say what they did (director, script-writer and so on). A scientific paper could list which co-authors carried out the critical experiments, which did the critical analyses and which ‘just’ arranged the funding.

Robert P. Bywater

Novo Nordisk A/S, Novo Nordisk Park,
DK-2760 Måløv, Denmark

Dr Bywater’s proposal has been *Nature*’s policy since 1999, as outlined in two Opinion articles (*Nature* 415, 819; 2002 and *Nature* 399, 393; 1999). *Nature* encourages co-authors to specify the contribution they made to the paper in the acknowledgements section.

These letters are a small selection of the many that *Nature* and Dr Lawrence have received about his Commentary. The overwhelming majority were in support of his views. — Editor, Correspondence.

Arab science is not stifled by censorship

Sir—Including my photograph in the box “Science veiled in secrecy” on page 122 of your News Feature on Arab science—“Blooms in the desert” (*Nature* 416, 120–122; 2002)—might suggest to your readers that I support the views expressed in the article, about censorship and secrecy in Arab countries.

In fact, I emphatically do not share any of these views. Given the level of cooperation between our organization and all leading Arab science and technology institutions, I do not wish in any way to be associated with statements such as those made in this box.

Omar Bizri

Chief, Technology Section, United Nations
Economic and Social Commission for Western Asia,
Beirut, Lebanon

Policy, politics and perspective

The scientific community must distinguish analysis from advocacy.

Roger A. Pielke Jr

Publication of Bjorn Lomborg's *The Skeptical Environmentalist* in September 2001 was immediately followed by an unprecedented mobilization of environmental advocates against the book, its author and publisher (see 'From teapot to tempest', below), the reverberations of which are still continuing. The Lomborg affair merits attention not because of its robust criticisms, character assassination and pressure politics — these are nothing new — but because its extremeness could mark a watershed in how science relates to policy and politics. Whether BSE, ecosystem functioning, genetically modified organisms (GMOs), cloning or vaccination, science is increasingly the battlefield on which political advocates, not to mention lawyers and those with commercial interests, manipulate 'facts' to support their positions. It is urgent that the scientific community changes if it is to prevent science's contribution to effective policy development from being diminished, and the practice of science from being compromised.

Gridlock

Politicization of science has always been, and always will be, integral to political advocacy. To take some examples:

Global climate change has been the subject of raging arguments for decades. Some people and organizations suggest that the problem is minimal and will take care of itself. Others advocate dramatic, immediate changes to global energy policies. Yet others believe that science cannot definitively predict climate change or economic futures and that a more sensible political approach would emphasize so-called 'no-regrets' adaptation and mitigation.

Nuclear power has been a subject of intense political debate and activism for even longer. Considerable scientific effort has been devoted to assessing risks associated with nuclear plants and nuclear-waste storage, with advocates and opponents of nuclear power each using 'science' to support their positions. Many other issues involving risk assessments (for example, GMOs and chemicals added to food and water) share similar characteristics.

Biodiversity is another controversy in which ecologists and the public are focused on science. Most of the attention is directed on the relationship of biodiversity to ecosystem functioning. How much diversity is desirable, and for which species? The argument is often less about science than about

different philosophical approaches to ecosystem management, which includes how different people value the aesthetic, spiritual and ethical aspects of biodiversity.

These and other issues are complicated by powerful economic and political interests that not only have high stakes associated with alternative policy outcomes, but also employ scientific experts to support their positions. Despite vigorous differences of opinion between combatants, without exception they share the belief that science is the appropriate battleground, as well as the assumption that if a perception can be created that science is on your side, you will win. In the Lomborg case, for example, *Scientific American's* editors subtitled the magazine's collection of responses "Science defends itself against *The Skeptical Environmentalist*", as if the authors are speaking 'for' science, rather than (as was actually the case) criticizing Lomborg's scientific claims and their significance for policy. It is this distinction, between science and policy advice, that the scientific community needs to address.

Political decisions that involve different interest groups are inherently difficult to make, because any adopted policy is bound to infringe on someone's (overt or vested) interests. The process of achieving a legitimate outcome involves bargaining, negotiation and compromise — political manoeuvring that is well beyond the scope of science.

'Science' is not a monolithic entity. In the words of the chemist Henry Bauer, it is "a mosaic of the beliefs of many little scientific

groups". This diversity stems from the perspectives of individual scientists themselves, as well as from the nature of the objects studied. In climate research, for example, the scientific uncertainties are so great that it is impossible to exclude a wide range of future outcomes, ranging from relatively mild to globally catastrophic. Even if science could provide a crystal-ball view of the future, justification for any particular climate policy would depend on more than what science alone is capable of providing, including desirable societal and environmental outcomes.

Some scientists believe that 'science' alone provides a sufficient basis for decision-making, in that a problem is identified, various hypotheses are tested, remedial policies suggested and implemented — then the situation improves. But putting the onus of problem resolution onto science brings all the messy realities of politics into the practice of science. Rather than making politics more scientific, this approach, in fact, makes science more political. Indeed, I have never come across any real-world policy issue involving science and decision-making that has resolved itself in this logical but oversimplistic manner.

Why science has become political

The answer lies in an 'iron triangle' of mutually reinforcing interests. In one corner is the politician loath to make a decision that will upset part of her constituency; she is consequently happy to pass the onus of resolution

From teapot to tempest

The Skeptical Environmentalist by statistician Bjorn Lomborg (Cambridge University Press, 2001) is a survey of global environmental problems and issues. The nub of the controversy is Lomborg's endorsement, from his self-described 'environmentalist's' perspective, of the work of the late Julian Simon, a renowned Copernican economist espoused by the US right wing for his optimism about environmental issues such as population growth. Lomborg's conclusion is that, broadly, "things are getting better" in a range of environmental areas.

To illustrate the points made in this Commentary, a criticism of Lomborg's science can be found at www.ucsusa.org/environment/mahlgan.pdf, and a criticism of the significance of Lomborg's argument is in *The Times Higher Education Supplement* 16 Nov. 2001, p 23.

For a selection of other reviews and comments about the controversy see:

Pimm, S. & Harvey, J. *Nature* **414**, 149 (2001), and Correspondence by Trewavas, A. *Nature* **414**, 581–582 (2001) and Budiansky, S. *Nature* **415**, 364 (2002).
Grubb, M. *Science* **294**, 1285–1287 (2001).
Schneider, S., Holdren, J. P., Bongaarts, J. & Lovejoy, T. (with editorial by J. Rennie). *Sci. Am.* **286**, 61–71; 2002.
The Economist 6 Sept. 2001, unsigned editorial and readers' letters in 14 Feb. and 28 Feb. 2002 issues.

Dutton, D. *Washington Post* 21 Oct. 2001; p BW01.
Ridley, M. *Spectator* 23 Feb. 2002; pp 10–11.
♦ tompaine.com/feature.cfm/ID/4791
♦ www.lomborg.com (Lomborg's own website)
♦ www.anti-lomborg.com
♦ www.gristmagazine.com/grist/books/lomborg121201.asp
Union of Concerned Scientists Commentary
♦ www.ucsusa.org/index.html

on to the scientist, typically by way of a generously funded government programme for research designed to provide 'answers'. These answers are unlikely to be provided until after the politician has moved on.

In another corner is the scientist, being offered resources to perform research not only to expand knowledge in the field, but to resolve important policy issues. Two birds with one stone! Hence the scientist accepts the generous resources for research, and with it, the mandate to provide 'answers'.

In the third corner is the advocate, looking for scientific data to provide a compelling justification for his political, societal, environmental or business goal. His opponent thinks along the same lines, of course. Science brings with it an air of impartiality and being 'above the fray' but, ironically, its use in such advocacy actually undermines impartiality. Similar to the politician, advocates (and their lawyers or consultants) look to science to definitively resolve political debate — so long as the resolution is in line with their preferences.

More and more, this mutually reinforcing iron triangle of shared interests is replacing explicit political debate about values and interests. 'Debate' over scientific issues increasingly relies on tactics including personal attack and criticizing processes (for example, peer review or sources of funding), through paid advertisements and other publicity campaigns. As political battles are waged through 'science', many scientists are willing to adopt tactics of demagoguery and character assassination as well as, or even instead of, reasoned argument — take, for example, the extremes of debates on genetically modified crops or global warming. Science is becoming yet another playing field for power politics, complete with the trappings of media spin and a win-at-all-costs attitude. Sadly, much of what science can offer policy-makers, and hence society, is being lost.

An alternative

Imagine a world in which scientific advice is provided to society only through established political affiliation, in which scientific journals are published through party structures — *Labour's Science* or *Republican Nature*. Public funding for research would be provided to political party organizations to disseminate as they wished, perhaps relying on traditional peer review, perhaps not. It would be difficult to find any practicing scientist (including myself) in favour of this vision, but the increasing politicization of science today could allow this to happen in practice.

To understand the politicization of science, it is essential to differentiate scientific results from the policy significance of those results. To illustrate the distinction, consider the Intergovernmental Panel on Climate Change (IPCC)'s conclusion that the global average temperature in 2100 will increase

Rather than making politics more scientific, this makes science more political

from 1.4 °C to 5.8 °C. Explaining what this scientific result means to the non-specialist may take some effort — it may require explaining the origins of the estimates, how 'global average' is defined, trends, conditions and the confidence levels of the projection. Yet, crucially, all this is different from an assessment of the significance of this conclusion for action ('policy'), which depends on how the results ('science') are related to valued outcomes, such as human health, environmental sustainability, economic prosperity and so on.

Assessing the significance of science for policy requires a clear distinction of policy analysis from political advocacy. The former increases the range of alternatives available to decision-makers by clearly associating scientific results with a range of choices and outcomes. The latter seeks to decrease the range of alternatives (often to a single desired outcome). Because scientific results typically have a degree of uncertainty, and because a range of alternatives can achieve particular policy outcomes, commitment to a particular policy involves considerations that go well beyond science.

Advocates can and frequently do provide valuable policy guidance. But to guard against the politicization of science, the independent scientific community must take responsibility for assessing the significance of scientific results for policy. A well-known example of such an attempt to provide independent scientific guidance is found in the IPCC, which has largely received positive reviews of its assessments of climate change (see *Nature* 412, 112; 2001). But the IPCC does not explicitly assess scientific results in the context of particular policies, which may be its greatest weakness. The IPCC only assesses knowledge of climate-change science, impacts and economics, and not their policy significance. Consequently, to understand the significance of the IPCC's analyses for alternative courses of action, a decision-maker is forced to rely almost exclusively on the interpretations (and misinterpretations) provided by corporations, government agencies or interest groups. Invariably, such interpretations are at odds with one another, yet consistent with all or parts of the IPCC's results. When well-intentioned IPCC scientists enter the political fray as individuals, the IPCC itself becomes politicized.

One solution in the IPCC case would be to establish a new, independent group on

policy, explicitly for assessing the significance of the scientific results in the context of policy. This kind of group could assess a broad range of alternative actions that are consistent with IPCC assessments without endorsing a particular alternative. (This group could also provide valuable feedback to the research community as to the issues that need more attention.)

To take another example, many granting agencies now ask in their evaluation criteria whether the proposed research will benefit society. Although many scientists support this principle, they do not have the expertise to assess their work in this way. Consequently, authoritative and non-partisan bodies such as national science academies and other independent scientific societies need to assume more responsibility for helping scientists place the significance of their research into a policy context. The US National Academy of Sciences occasionally provides such guidance on a range of issues, for example arsenic in water, reproductive cloning and streamflow for salmon and farmers, but, not surprisingly, has a penchant for recommending 'more research' as the preferred policy alternative in almost every context!

Recent effort to revive the US Congress Office of Technology Assessment (OTA), terminated by Congress in 1995, recognizes a void. Whether or not the effort succeeds, it has helped to focus discussion on the two-way connections of science and policy in the political arena. The OTA, for the most part, avoided partisanship by associating scientific and technological results with a wide range of possible policy outcomes, leaving for decision-makers the task of selecting particular courses of action. The scientific community itself must recognize its own interests and systematically provide guidance on the significance of scientific results, using some such mechanism.

Political advocates will always selectively use and misuse scientific data to support their agendas. Today's scientists need to understand the consequences for science of relying on political advocacy as the primary mechanism of connecting science with policy. There is little sense in yearning for a bygone era when science was construed as 'value-free'. Rather, the scientific community should consider providing insight in a more systematic way through independent, authoritative bodies, so that the choices available to policy-makers and the public are expanded. In some cases this may help to find a way through gridlock and political stalemate, in others it may offer a realistic view about the limits of using science in policy and politics. ■

Roger A. Pielke Jr is at the Center for Science and Technology Policy Research, University of Colorado/CIRES, 1333 Grandview Avenue, Boulder, Colorado 80309-0488, USA.

What the Nahua knew

The secrets of Mexico's plants had been revealed by the sixteenth century.

The Mexican Treasury: The Writings of Dr. Francisco Hernández

edited by Simon Varey. Translated by Rafael Chabrán, Cynthia L. Chamberlin & Simon Varey

Stanford University Press: 2001. 281 pp. £40, \$65

Searching for the Secrets of Nature: The Life and Works of Dr. Francisco Hernández

edited by Simon Varey, Rafael Chabrán & Dora B. Weiner

Stanford University Press: 2001. 229 pp. £40, \$65

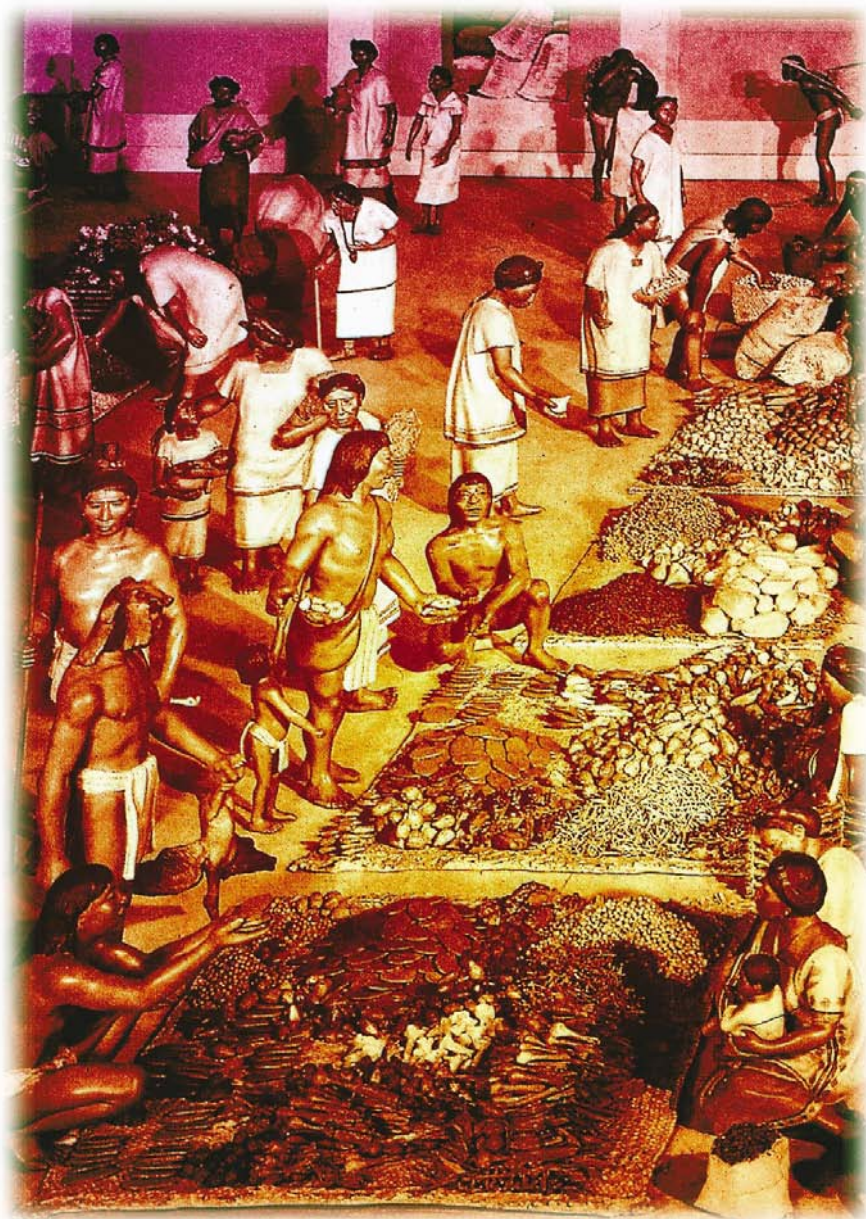
Karen Reeds

In 1570, Philip II of Spain sent Francisco Hernández to Mexico on the most ambitious government-sponsored scientific expedition since antiquity. The royal physician was ordered to survey the plants, animals and minerals of New Spain, to gather and test information about their uses from "doctors, medicine men, herbalists, Indians, and other persons with knowledge in such matters," and to describe and illustrate the natural history, geography and "antiquities" of the region.

Despite budget cuts, unfamiliar languages, hazardous travel and illness, the seven years Hernández spent in America were tremendously productive. He presented Philip with six folio volumes of text and ten more of paintings commissioned from native artists, describing more than 3,000 plants (at the time, European botanists knew about 6,000 plants), 400 animals and 35 minerals. Ten years later, Hernández died, heartbroken because his *Natural History of New Spain* had not made it into print.

What happened? Other Renaissance botanists' plans for monumental herbals had been scuttled by the high cost of producing the illustrations, and Philip's treasury went bankrupt while Hernández was in Mexico. Yet, when just about any book on New World exploration captured the European imagination, the *Natural History of New Spain* would have been a certain best-seller. An enterprising printer could easily have raised the necessary capital.

But there were other obstacles to publication. Philip had little incentive to subsidize the release of such valuable information about New Spain's resources. The few contemporaries who saw Hernández's early draft were baffled by



This reconstruction of a spice market shows the diversity of plants available to the Nahua.

his decision to organize it by the unfamiliar categories of plant classification, medical uses and names used by the Nahua, the dominant population group of central Mexico. In *Searching for the Secrets of Nature*, Carmen Benito-Vessels argues that Hernández's religious orthodoxy was suspect. His friends (and quite possibly his family) included crypto-Jews and Erasmianists; and, by naming the creatures of Mexico, he encroached on God's prerogative. The Inquisition's censors would never have passed the book.

Instead, the contents of Hernández's manuscripts trickled out to the world over

the next three centuries — a fate charted exhaustively through these two volumes' excerpts, essays, chronologies, and, most revealingly, in Chabrán and Varey's tangled stemma diagram in *The Mexican Treasury*. The best-known version of Hernández's text and illustrations, *Rerum medicarum Novae Hispaniae thesaurus*, was based on a digest made in 1580 on Philip's orders. That compilation landed up in Italy (where Galileo admired it); after decades of dithering, it was published by the Accademia Lincei in 1651. Hernández's presentation manuscript to Philip was destroyed by fire in 1671. The early draft survived "doing battle

with cockroaches and worms" in a Madrid library; only parts have been edited.

Suppose Hernández had published his work soon after his return to Spain in 1577. What difference would it have made to the course of science? *The Natural History of New Spain* would, I believe, have ranked with Leonhart Fuchs's great herbal of 1542 and Vesalius's anatomy of 1543 as a pivotal work of Renaissance medicine and biology. It would have challenged botanists to come up with classification schemes and nomenclature to encompass so much new data. It would have provided other monarchs and expeditions with a model for scientific exploration, and it would have given both Spain and Mexico a reputation for scientific originality.

Its publication could have had even wider consequences for world history if its readers had absorbed Hernández's respect for his Mexican colleagues, their culture, and their medical skill. Fearing, with good reason, that epidemics would wipe out Mexico's expert healers, Hernández took pains to translate his work into Náhuatl. In his will he left a substantial bequest to pay all those Indian herb-gatherers, doctors and painters who had helped him, "if His Majesty does not make recompense to them".

Although these two new volumes can be bought separately, it would be absurd to do so. They were conceived as a single project and are admirably cross-referenced by the editors. *The Mexican Treasury* provides the first careful documentation in any language of the dissemination of Hernández's discoveries and translates a tantalizing sample of Hernández's texts, drawing both from surviving manuscripts and from the Spanish, Mexican, Italian, Dutch and English books on natural history that mined his work. The 18 essays in *Searching for the Secrets of Nature* are the first thorough discussion of Hernández's life, milieu and influence to appear in English. Together, these books represent a remarkable scholarly achievement and a splendid tribute to a remarkable man who linked two rich civilizations.

But these books are only a beginning. Some of the medicinal plants Hernández encountered are still used in Mexican communities. If I ran the US National Institutes of Health or a major pharmaceutical company, I'd invest a small fraction of my research budget in a complete edition, translation and ethnobotanical study of Hernández's surviving texts and illustrations. It is not too late to find out what the Nahua knew. ■

Karen Reeds is affiliated with the Princeton Research Forum/National Coalition of Independent Scholars, 301 North Harrison Street, Princeton, New Jersey 08540, USA. She is the author of *Botany in Medieval and Renaissance Universities*.



RONALD GRANT ARCHIVE

Burning issue: the 1982 film *Quest For Fire* argues that trade may have influenced the early use of fire.

Hominid economics

Second Nature: Economic Origins of Human Evolution

By Haim Ofek

Cambridge University Press: 2001. 264 pp.

£45, \$74.95 (hbk); £17.95, \$27.95 (pbk)

Christopher Wills

The literature of palaeontology is replete with attempts to explain why the hominid lineage took the evolutionary direction that led to *Homo sapiens*. The problem has several parts. First, what was the initial triggering event? Was it upright posture, climatic change or perhaps a high frequency of a gene that influences curiosity that led one small group of our ancestors out of the forest and into the more open savannah? Whatever it was, it must have happened more than four million years ago.

Second, what caused the approximately simultaneous appearance, about two million years ago, of larger brains, more sensitive hands, stone tools and the ability to travel across entire continents? Third, what triggered the Great Leap Forward that led to representational art and sophisticated tools? This leap, now realized to have been more of a sustained, hesitant shuffle, took place over the period from 300,000 to 40,000 years ago, first in Africa and later more widely across the Old World. And finally, what was the cause of the agricultural revolution that started around 10,000 years ago and that eventually led to cities and the whole panoply of the modern world?

Haim Ofek, an economist at Binghamton University in New York, argues in this provocative book that economic factors

played an important and under-appreciated role in all four of these transitions. The earliest role for economics must have been in the form of a division of labour in the first hominid societies, powered by the need for the simultaneous optimization of foraging strategy, defence, hunting and tool-making.

Ofek argues that economic theory can also help to explain how the ability to use fire has changed with time. During the many millennia before the technology was developed that made it easy to start fires, early tribes might have maintained an 'incendiary hub' — a few fires are kept going by individual effort and are periodically transplanted to members of the group that have let their own fires go out. This might have led to cooperation and communication. (For a different view of the early history of fire, which emphasizes the benefits of trade and exploration, I recommend the marvellous 1982 film directed by Jean-Jacques Annaud, *Quest for Fire*.)

Economic connection to later events such as the agricultural revolution are easier to make because the evidence is more plentiful. The author makes a strong case, as others have done before him, that this revolution in behaviour resulted from a combination of stable climate, geography and an increase in trade between early agriculturalists and pastoralists. Ofek usefully cites such disparate sources as the political economist Adam Smith, Pliny the Elder and the nineteenth-century agricultural economist Johann Heinrich von Thünen, figures who are not often quoted in studies of evolution.

Ofek makes several interesting connections between economics and biology, but he often wanders from the point, and fails to demonstrate clearly how economic pressures might have contributed to the evolutionary process itself. In particular, he states

that the division of labour in human societies cannot have a genetic basis, and so cannot be subject to evolutionary pressures. He contrasts humans to the social insects, in which there has been selection for division of labour. Although he does not explore the biology, it is clear that these insects carry genes that are expressed in different ways under different environmental conditions. So all the female members of the population, regardless of their genotype, are capable of developing into queens, workers or soldiers.

Such a situation, of course, is not found in humans. But Ofek ignores the possibility that polymorphism for genes that influence behaviour in hominid populations could have been maintained, and also increased, by a combination of mutation and frequency-dependent selection. In our species, concert pianists, master mechanics, soldiers, nurses and economists, who are all the result of a combination of genes and environment, can all contribute to society. One of the exciting prospects for future research is that we may soon find some of these polymorphic genes that influence behaviour. When we find them, we can explore the selective pressures that have maintained these genetic polymorphisms in our remarkable and talented species. Some of these pressures will be found to be economic, but many will not. ■

Christopher Wills is in the Division of Biology, University of California at San Diego, San Diego, California 92093-0116, USA.

All the science that's fit to print, and more

Doing Science: Design, Analysis, and Communication of Scientific Research

by Ivan Valiela

Oxford University Press: 2001. 304 pp. \$75, £55 (hbk); \$39.50, £35 (pbk)

David Ward

Have you ever thought it would be nice to have a book that contains all the generic information that a student will need to know in order to design, analyse and communicate the results of scientific research? I certainly have. Ivan Valiela set out to write such a book, and *Doing Science* is the result. He gives us a very brief introduction to elementary philosophical issues in science, the most cursory of glances at hypothesis testing, writes a little about experimental design, provides a thorough discussion of scientific writing and the graphical presentation of scientific results, and ends, rather oddly, with an afterthought on perceptions and criticisms of science. *Doing Science* is well written and peppered with interesting historical asides and pertinent examples from Valiela's own

research field, marine biology. He has created an interesting, readable book out of a subject that he admits is potentially "dreadfully dull".

Could a student pick up this book and use it as a road map to do scientific research? Not really. The section on experimental design is passable for simple scenarios but the section on hypothesis testing is too elementary and is inconsistent enough to get you into trouble rather than out of it. Many statistical issues are covered in such a way that students with no background in statistics will be left floundering, whereas those with some prior knowledge will find it little more than a quick reminder of a particular technique they seldom used. For those of you intending to use Valiela's examples in the statistical section, beware of errors in some of the tables (such as inconsistent use of the terms 'variance' and 'variation', which are occasionally used interchangeably). Furthermore, although the title implies that the book is a general guide to doing science, my feeling is that 'Doing Biology' would have been more appropriate, considering the emphases in the design and analysis sections of the book.

The above-mentioned quibbles aside, I found the scientific writing section to be superb, and the advice on the graphical display of data to be outstanding, albeit with liberal borrowings from E. R. Tufte's seminal work *The Visual Display of Quantitative Information* 2nd edn, (Graphics Press, 2001). These tips on writing and graphical presentation are not entirely new, but it is nice to see so much under one roof and so engagingly presented. And these sections are

general enough to be suitable for biologists and non-biologists alike.

My greatest concern about this book is its inconsistency. I think the design and analysis sections would be appropriate for the introductory level of a bachelor's degree, although the student will need considerable help from a more substantial statistical text; I find J. H. Zar's *Biostatistical Analysis* (Prentice-Hall, 1999) to be one of the more readable of these. The writing and presentation sections of *Doing Science* are more suitable for senior undergraduates or even graduate students. It is too inconsistent in detail and emphasis to be suitable as a course text, but I would strongly recommend it as a general reference book.

The book's easy style will make it a favourite of students who are looking for a quick introduction to a range of topics in this field. I think that it will be particularly useful for students whose first language is not English and for those who need to write a scientific paper in English for the first time. Indeed, Valiela is particularly sensitive to this issue and goes to great lengths to explain his views on the communication of science in English. Students from universities where there is still (unfortunately, I might add) a great dichotomy between fact-based teaching at the undergraduate level and discovery-based teaching and/or research at the graduate level (or who have switched between universities with these traditions) will find this text a useful way to get up to speed at the beginning of a graduate career. ■

David Ward is in the Department of Conservation Ecology, University of Stellenbosch, Matieland 7602, South Africa.

Floral fortunes

The saffron crocus, *Crocus sativus* (shown here), was at the heart of the dye industry in Britain in the Middle Ages, and gave its name to the town of Saffron Walden in Essex. But the flower's popularity gradually declined as foreign crocuses with wider ranges of colour were introduced into the country. The tale is typical of those told in *Flora: The Illustrated History of the Garden Flower* by Brent Elliott (Scriptum Editions, £45), which draws on the extensive archive of the Royal Horticultural Society. The book charts the history of Britain's wealth of garden flowers as for hundreds of years new specimens have been imported from around the world.



Unearthing the gems

Mikhail V. Blagosklonny and
Arthur B. Pardee

At this opportune time, conceptual biology is being born. Millions of easily retrievable facts are being accumulated in databases, from a variety of sources in seemingly unrelated fields, and from thousands of journals. New knowledge can be generated by 'reviewing' these accumulated results in a concept-driven manner, linking them into testable chains and networks.

Molecular biology has moved from an era of data collection into one of hypothesis-driven, experimental research. By connecting several facts, a hypothesis can be formulated in terms that are testable by experiments. Yet many key experiments are already in the literature, and a search would have revealed them.

As the biochemists Douglas Hanahan and Robert Weinberg have emphasized, exponentially increasing amounts of information add "further layers of complexity to a scientific literature that is already complex almost beyond measure". This complexity and overproduction of data can be an obstacle to efficient research unless data are conceptually organized. But even more importantly, this accumulation of data provides new opportunities: when searchable results are readily available (for example, in databases such as Medline and PubMed), the data themselves provide the domain for investigation. Many 'future' predictions have already been tested using these enormous data collections, through related experiments that were carried out for other reasons, and which 'only' need to be revealed in the new context.

Connecting separate facts into new concepts is analogous to combining the 26 letters of the alphabet into languages. One can generate enormous diversity without inventing new letters. These concepts (words), in turn, constitute pieces of more complex concepts (sentences, paragraphs, chapters, books). We call this process 'conceptual' research, to

distinguish it from automated data-mining and from conventional theoretical biology (already defined as a separate discipline that must be integrated into biology, as discussed by Dennis Bray in an earlier essay in this series).

Conceptual biology, as we see it, is not a distinct type of science, but rather it has a different source: the information in databases. For example, conceptual oncology is a branch of oncology, not of theoretical biology. By logical, critical analysis of existing facts and models, one can generate a hypothesis in which predictions are formulated in testable terms, and then search for relevant information among published reports of experiments that may have had a different purpose altogether.

In general, the more publications there are in a field, the more opportunities that field presents for conceptual research. The p53 field, for instance, is a goldmine. A Medline search identifies 22,749 publications on this tumour-suppressor protein. In 1990, all the necessary data for the concept of feedback control of p53 function were available, although this function was recognized only recently. Unlike wild-type p53, mutant p53 is a stable protein, and mutant p53 cannot *trans*-activate. By linking these two facts, the concept that loss of function causes p53 stabilization predicts the existence of the 'feedback-control protein' Mdm2. This link also predicts that the viral SV40 oncoprotein will cause p53 to be stabilized. Instead of experimentally testing this prediction, one can discover that the answer is not only known, but that this stabilization actually led to the discovery of p53 in 1979.

Enormous collections of data allow hypotheses to be generated and tested using pre-existing data. Many experimental discoveries arise from chance observations. Similarly, data-searching can reveal unexpected connections. Thus, by searching successive pairs of terms, a chain or network of connections can be generated. For example, inputting the search term 'kinase C + apoptosis' generates about 4,000 citations; 'kinase C + NF- κ B' yields about 1,000; and 'NF- κ B + apoptosis' gives around 1,000, suggesting that NF- κ B might be an intermediary between kinase C and apoptosis. Similarly, the 100 citations that include 'kinase C + I κ B kinase (I κ K)' suggest that I κ K is also an intermediary. Furthermore, kinase C is linked to the drug Go6976 by 100 citations. These combined findings suggest that Go6976 might activate apoptosis by inhibiting kinase C, thereby blocking activation of I κ K and NF- κ B.

Apparently, one is not licensed to theorize without providing new data. But this is a sociological problem, not a scientific one, as Bray points out. For a theory to be valid, it

Conceptual biology

The conceptual review should take its place as an essential component of scientific research.

does not matter on whose results it is based, as long as they are correct. Published results can be preferable for this purpose; they tend to be less error-prone and more accurate because of peer review, and because they are more often tested and built upon by others.

Elemental, conceptual review articles, in the form of opinions, perspectives, 'trends' and so on, are on the rise. Yet some believe that reviews inherently lack novelty because they tap data that are already in the literature. Reviews are often underestimated or even scorned by experimental researchers, and because they are not regarded as original in themselves, have a smaller impact on science.

Can a review provide new knowledge? A review can constitute a comprehensive summary of the data in the field — this type of writing educates but does not directly generate new knowledge. But a 'conceptual' review, on the other hand, can generate knowledge by revealing 'cryptic' data and testing hypotheses by published experiments.

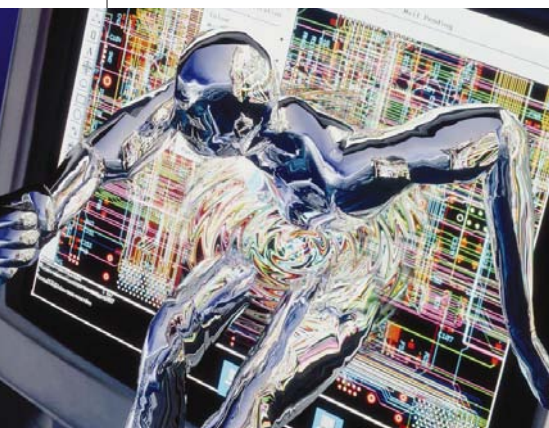
Conceptual biology should be recognized and criteria established for its publications — new, testable conclusions, supported by published data. In biological systems, everything is interconnected, and ostensibly unrelated fields are related — the separation of biology into different disciplines is artificial. Conceptual research can encompass many fields without limitation. In comparison with labour-based research, conceptual research is more cost-effective; indeed, verification of a hypothesis using existing data does not limit research to scientists in well-resourced fields or countries. Hypothesis-driven, experimental research will continue to be a cornerstone of biology, but it should strike up a partnership with the essential components of theoretical and conceptual research. ■

Mikhail V. Blagosklonny is at the Brander Cancer Research Institute, New York Medical College, Hawthorne, New York 10523, USA.

Arthur Pardee is at the Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

FURTHER READING

Weinberg, R. *Trends Biochem. Sci.* **26**, 207–208 (2001).
Blagosklonny, M. V. *Oncogene* **15**, 1889–1893 (1997).
Goodman, A. B. *Proc. Natl Acad. Sci. USA* **95**, 7240–7244 (1998).
Hanahan, D. & Weinberg, R. A. *Cell* **100**, 57–70 (2000).
Bray, D. *Nature* **412**, 863 (2001).



Evolution of suicidal signals

Mike Speed and Graeme D. Ruxton

To deter predators, some organisms have evolved protective mechanisms and warning colorations. Innovative mathematical modelling shows how the first animals with these signals might have managed to survive.

Many animals have warning coloration to show that they are best avoided by predators. How such signals initially evolve is a mystery that has entertained biologists for some years. Thomas Sherratt, writing in *Proceedings of the Royal Society*¹, has provided a new and plausible way of tackling the matter.

Almost all elementary textbooks in behavioural ecology present the idea that an animal with a good anti-predator defence should advertise its unprofitability, in order to escape the costs of attacks by predators that have learned (through previous unpleasant experiences) to link the warning signal to unpalatability. This is the explanation for the existence of the conspicuous colorations in toxic or spiky insects (Fig. 1), in venomous snakes and frogs, and perhaps even in some birds².

Dig a little deeper, however, and we uncover a famous evolutionary quandary^{3–5}. A warning signal provides protection from predators that have learned to avoid such individuals, but that same signal is likely to make its bearer more obvious to those 'naïve' predators that have yet to learn. Possessing a warning signal makes sense if the signal is common, because the cost of educating naïve predators is shared among many individuals. But when they originate, new warning signals will be rare and, it seems, their owners should be rapidly snuffed out by naïve predators. So the initial evolution of warning signals — making seemingly 'suicidal prey' — presents a serious challenge for evolutionary biology.

Sherratt proposes a mathematical model that allows predators to co-evolve with their prey. Although there have been many attempts to model the evolution of warning signals in the past, this latest attempt is fundamentally different in that it explicitly acknowledges that the decisions made by one predator can help other predators decide which prey are worth attacking and which should be avoided. Sherratt envisages a situation with prey that vary in both their palatability and their conspicuousness. These assumptions match our observations of the world around us, and are essential for his model to work well. If edible prey have lower chances of surviving an attack by a predator than 'defended' individuals, predators will act to filter out the edible prey. When predators act as filters in this way, they soon remove the conspicuous edible individuals,



Figure 1 Unpalatable advantage. Many insects, such as this sharp-spined *Hyalophora cecropia* caterpillar, defend themselves against attack — and make that clear to predators through conspicuous colouring.

leaving a world in which conspicuous prey are reliably defended and unprofitable to eat, and less obvious prey are mainly edible. After each round of filtering, conspicuousness becomes an increasingly good predictor of unprofitability; in response to this, predators evolve greater wariness of conspicuous prey.

Other researchers⁴ have proposed that conspicuousness and predator wariness might become entwined in a runaway co-evolutionary process. But Sherratt formally shows that such a process can indeed occur — one in which, as conspicuousness becomes an increasingly reliable signal of unprofitability, more predators will increase their wariness of novel, detectable prey. So it becomes easier for a diversity of conspicuous warning signals to evolve and populate the world.

An intriguing feature of Sherratt's model is that various sets of circumstances can produce an evolutionarily stable combination of predatory strategies, in which some predators attack novel prey, some attack with caution (treating conspicuousness as a 'go slow' signal⁶) and others do not attack at all. This feature of the model may explain the surprising demonstration⁷ that wild birds in the same breeding population can show huge variability in their willingness to sample and incorporate new types of prey into their diets.

Have we at last got an explanation for the paradoxical evolution of suicidal prey? The answer seems to be 'yes and no'. Sherratt has provided a highly stimulating view of the evolution of warning signals, but at the same time he has opened up some quite profound

questions about the conditions that precede these signals. Is the correct starting point from which to consider the co-evolution of predators and prey a world in which defended and edible prey come in a wide diversity of forms (as Sherratt assumes in this model)? Or is it one that is overwhelmingly and monotonously drab, with some conspicuous mutants arising only rarely? If the second possibility is more likely, then further work will have to explore whether Sherratt's co-evolution theory produces the same result from this quite different starting place. Finally, we still need to explain why there was a bank of toxic, well-defended, inconspicuous prey from which individuals with warning coloration could arise in the first place. After all, common sense suggests that, to be worth the cost, toxicity needs to be advertised. ■

Mike Speed is in the Department of Biology, Liverpool Hope, Hope Park, Liverpool L16 9TJ, UK. e-mail: speedm@hope.ac.uk

Graeme D. Ruxton is in the Division of Environmental and Evolutionary Biology, Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, UK. e-mail: gdr1m@udcf.gla.ac.uk

1. Sherratt, T. N. *Proc. R. Soc. Lond. B* **269**, 741–746 (2002).
2. Dumbacher, J. P. & Fleischer, R. C. *Proc. R. Soc. Lond. B* **268**, 1971–1976 (2001).
3. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
4. Turner, J. R. G. *Symp. R. Ent. Soc. Lond.* **11**, 141–161 (1984).
5. Guilford, T. in *Insect Defences: Strategies of Prey and Predators* (eds Evans, D. L. & Schmidt, J. O.) 23–61 (State Univ. New York Press, New York, 1990).
6. Guilford, T. *J. Theor. Biol.* **170**, 311–316 (1994).
7. Marples, N. M., Roper, T. J. & Harper, D. G. C. *Oikos* **83**, 161–165 (1998).

Physical chemistry

Sculpting ice out of water

Srikanth Sastry

Computer simulations are illuminating the molecular processes through which water is transformed into ice — and offering insight into crystallization more generally.

When water becomes cold enough, it freezes into ice. Capturing this everyday process on a computer, however, is far from routine. Although there have been simulations of water crystallization, researchers have usually managed this only by creating artificial conditions — for example, by confining the water in a restricted space¹. In such studies, including a recent one in which there was spontaneous crystallization², the resulting crystallized state was not the usual hexagonal ice that forms at atmospheric pressure.

On page 409 of this issue, Matsumoto, Saito and Ohmine³ report the first successful simulation of the formation of hexagonal ice. Their results open the way to understanding the complex kinetics of how water freezes, at the microscopic level. Understanding the kinetics of such phase transformations is important in many areas, from designing new materials⁴ to determining protein structure by crystallography⁵.

The freezing of water into ice is the most familiar example of a first-order phase tran-

sition — the transformation of matter from one state to another that brings about a marked change in properties. When water is cooled below the freezing point, 0 °C at ambient pressure, the regular structure of ice, with strong attractive interactions between molecules, becomes the thermodynamically preferred state. Nevertheless, it is relatively easy to cool water below 0 °C without the liquid crystallizing. Despite the fact that, in this supercooled state, liquid water is thermodynamically less stable than ice, it remains stable as long as there are no large changes or fluctuations in the system: if a perturbation exceeds a certain threshold, the liquid will transform into ice.

The threshold to crystallization is defined by the formation of a 'critical nucleus', which is expected to be a crystallite of ice that is large enough not to dissolve back into the liquid, and which will grow until the entire sample is transformed into ice. Agents that can cause nucleation include dirt particles, around which a nucleus forms (this is how artificial rain is seeded), or mechanical

agitation (shaking the container). Barring such external help, nucleation must occur spontaneously: ever-present thermal fluctuations can occasionally transform local regions of the liquid from liquid-like structure to a more crystal-like configuration. When such fluctuations cross the critical threshold, nucleation and growth of ice ensue.

Matsumoto *et al.*³ have tried to work out the details, at the molecular level, of how the critical nucleus forms and grows in water. They simulated the dynamics of water molecules, using knowledge of the forces between the molecules, to trace the evolution of molecular arrangements over microsecond timescales, which are long by present-day standards of computer simulation. And this is where the difficulty arises. Matsumoto *et al.* needed to generate many microsecond-long time histories (or trajectories) in order to find even one trajectory that led to crystallization. The reason for the difficulty in finding the crystal state, they say, is the complexity of water's potential-energy surface — this is the landscape-like map that describes how the total energy of the collection of water molecules varies as the individual molecules move.

Each configuration of molecules corresponds to a specific value of potential energy for the system. Configurations can be grouped together such that the potential energy varies smoothly between them, and a path connecting them exists along which there will be no potential-energy barrier.

Materials science

Repellent behaviour

Rain and condensation can make it difficult to see through spectacles or vehicle windscreens. One solution to this everyday problem lies in coating the optical surface with an ultra-water-repellent, transparent film. But for manufacturing purposes, the coatings on plastic surfaces (for example, lightweight spectacle lenses) need to be produced at temperatures below 100 °C to avoid damaging the surface.

Writing in *Chemical Vapor Deposition* (8, 47–50; 2002), Yuning Wu *et al.* report a strategy for producing ultra-water-repellent coatings that can be carried out at room temperature. They used a technique called microwave plasma-enhanced chemical-vapour deposition to deposit films of an organic monomer, trimethylmethoxysilane (TMMOS), at a partial pressure of 50 Pa, on a glass or plastic surface. The resulting

coatings are highly transparent and give the required high contact angle (above 150°) between the water and the coating (upper figure), and so a less wettable surface.

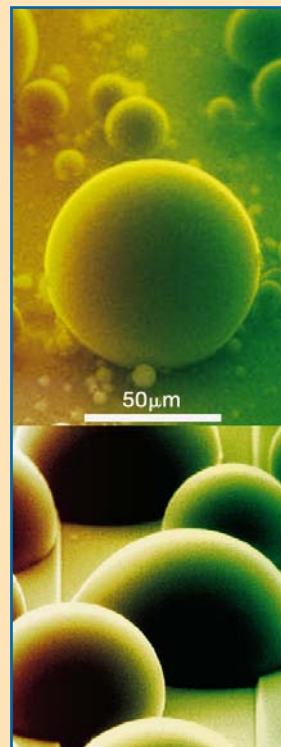
What is the cause of this ultra-water-repellence? The wettability of surfaces is governed by surface energy, chemistry and texture. The surface of the TMMOS film is believed to be terminated with methyl groups. So Wu *et al.* compared their coating with a chemically similar film, methyl-terminated self-assembled monolayers. These turned out to be much more wettable, with a water contact angle of about 100° (lower figure). A film of TMMOS prepared at a lower partial pressure (18 Pa) also resulted in a lower water contact angle (110°).

The answer to the ultra-water-repellent behaviour of TMMOS therefore appears to lie instead in its surface texture. Micrographs of the

coatings deposited at 50 Pa revealed non-wettable columns of the deposited film a few hundred nanometres in diameter, with pores in between. The authors suggest that the film created at 18 Pa was less water-repellent because its molecules are more densely packed and there are no pores. The methyl-terminated self-assembled monolayer appeared to be perfectly smooth.

The porous nature of the films deposited at 50 Pa means that the surface consists of both air and TMMOS. Wu *et al.* propose that the apparent surface energy of the film is therefore reduced, increasing its non-wettable properties. The authors don't expand on their theory to consider the events at the molecular scale. But the performance of the coatings shows that they could be of real use to spectacle-wearers out in the rain, or coming in from the cold.

Rosamund Daw



(An example would be a pair of configurations that differ only by a small twisting of one molecule with respect to its neighbours.) To simplify things further, each group of configurations can be represented by the configuration possessing the lowest energy in the group (that is, the local energy minimum). The dynamics can then be viewed as the system hopping from one minimum to another, requiring an intermediate increase in energy to cross each potential-energy barrier. These barriers, and the complexity of the pathway to crystalline structure, which defines the minimum with the lowest energy, determine the rate of crystallization.

Matsumoto *et al.*³ contend that, in contrast to simple atomic liquids such as argon, the pathway leading from the liquid to the lowest energy — crystalline — state is not smooth in the case of water. This idea is supported by an analysis of small clusters of water molecules by Wales *et al.*⁶. The explanation may lie in the fact that an open tetrahedral structure for the water molecules is energetically preferred, and that the hydrogen bonds between oxygen atoms sharing a common hydrogen atom are relatively flexible. So water can take on a globally disordered structure, but still retain the benefits of energetically favourable local arrangements.

A remarkable observation by Matsumoto *et al.* is that the critical nucleus that causes crystallization is itself not crystalline. In fact, there are other instances where the critical nucleus does not have the same structure as the equilibrium phase it nucleates (for example, the critical nucleus for a model liquid that will form a stable face-centred cubic crystal is a body-centred cubic crystallite⁷). And in modelling protein crystallization, ten Wolde and Frenkel⁵ showed that the initial nucleus is liquid-like when crystal nucleation takes place near a metastable critical point — in the phase diagram of the system they study, the liquid–gas critical point is pushed below the line at which the fluid is transformed to the solid, and therefore occurs in metastable conditions. Indeed, a metastable liquid–liquid critical point has been proposed for water⁸, so this idea has some appeal.

But there are potentially significant differences between the cases for water and proteins: the difference in density between the parent phase and the crystal that is nucleated is smaller in water; and the metastable critical point in water, if present, must be at much lower temperatures and higher pressures⁹. A more useful comparison in this context may be with liquid silicon¹⁰, for which rapid crystal nucleation occurs in conditions in the vicinity of a weak, first-order transition. Although it shares the loose-packed tetrahedral structure of water, silicon's bonding network is much less flexible. Sorting out the interplay of various factors in these systems would be a productive exercise.

To progress beyond the results of Matsumoto *et al.*, we need to study crystallization over a wider range of temperatures, and also to exploit methods⁷ for directly evaluating the free-energy barrier to nucleation. The computational effort required for the first task is formidable, while the second approach requires the identification of a small number of variables, or order parameters, for describing the progress of ice nucleation reliably. Matsumoto and colleagues' report on this point is discouraging — their attempts to identify useful order parameters for water have so far been unsuccessful.

The example of protein crystallization⁵ shows that a systematic approach to a complex question — the possibility of, and optimal strategy for, crystallizing proteins — is closely related to the seemingly much simpler question of how water freezes. Understanding the behaviour of water could help us to tackle other questions, such as when a liquid readily forms a disordered glass state¹¹ — an important issue in designing non-crystalline materials. Water may be a better starting point for addressing this

question than are atomic liquids. Given the significance of the general problem of crystal nucleation and growth, and the richness that the example of water brings to the question, work that builds on the results of Matsumoto *et al.*³ is bound to be rewarding. ■

Srikanth Sastry is in the Theoretical Sciences Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur Campus, Bangalore 560064, India. e-mail: sastry@jncasr.ac.in

1. Koga, K., Tanaka, H. & Zeng, X. C. *Nature* **408**, 564–567 (2000).
2. Yamada, M., Mossa, S., Stanley, H. E. & Sciortino, F. *Phys. Rev. Lett.* (in the press); <http://arXiv.org/abs/cond-mat/0202094> (2002).
3. Matsumoto, M., Saito, S. & Ohmine, I. *Nature* **416**, 409–413 (2002).
4. Joannopoulos, J. D. *Nature* **414**, 257–258 (2001).
5. ten Wolde, P. R. & Frenkel, D. *Science* **277**, 1975–1978 (1997).
6. Wales, D. J., Doye, J. P. K., Miller, M. A., Mortenson, P. N. & Walsh, T. R. *Adv. Chem. Phys.* **115**, 1–111 (2000).
7. ten Wolde, P. R., Ruiz-Montero, M. J. & Frenkel, D. *Phys. Rev. Lett.* **75**, 2714–2717 (1995).
8. Poole, P. H., Sciortino, F., Essmann, U. & Stanley, H. E. *Nature* **360**, 324–327 (1992).
9. Sciortino, F., Poole, P. H., Essmann, U. & Stanley, H. E. *Phys. Rev. E* **55**, 727–737 (1997).
10. Thompson, M. O. *et al. Phys. Rev. Lett.* **52**, 2360–2363 (1984).
11. Angell, C. A. in *Proc. Int. School of Physics, "Enrico Fermi" Course CXXXIV* (eds Mallamace, F. & Stanley, H. E.) 571 (IOS, Amsterdam, 1997).

Neurobiology

Serotonin sustains serenity

Solomon H. Snyder

An elegant variation on conventional gene-knockout techniques can delete a gene at specific times and locations in mice. The approach shows when and where a serotonin receptor protein is needed during development.

On page 396 of this issue, Gross and colleagues¹ look in depth at how serotonin, one of the chemical messengers that nerve cells use to communicate, is involved in anxiety. Perhaps one of the best known of these messengers, or neurotransmitters, serotonin has a role in many different neurobiological processes. For example, it helps to regulate our moods — a fact that has been well established since the 1950s, with the discovery that drugs that deplete serotonin precipitate depression whereas increasing serotonin levels has antidepressant effects. The idea that serotonin might also affect anxiety was first suspected in the 1980s following the serendipitous finding that buspirone, a drug developed to treat psychotic patients, is also useful for treating anxiety disorders, and stimulates a type of serotonin-detecting molecule in the body, the serotonin_{1A} receptor. Later came the discovery that mice that have been genetically engineered to lack this receptor, and so cannot respond normally to serotonin, show increased 'anxiety-like' behaviour^{2–4}.

But the underlying mechanisms have been elusive. For instance, the relevant brain regions have not been delineated. Moreover, the findings in receptor-deficient mice

appear to contradict observations that compounds that block serotonin_{1A} receptors do not cause anxiety in adult mice. Gross *et al.*¹ substantially clarify these issues. By using mice in which the serotonin_{1A} receptor can be knocked out at will, the authors show that the absence of the receptor in newborns does indeed lead to anxiety-like behaviour, whereas its knockout during adult life has no effect. Gross *et al.* also discriminate between the role of the receptors in the hindbrain and in forebrain structures such as the hippocampus and cerebral cortex.

Conventional gene-knockout techniques are powerful tools for working out what a protein does. But they have major limitations compared with using drugs (which might, for example, activate or inhibit the protein of interest). Genes tend to be knocked out during embryonic life, generally affecting the whole organism throughout its lifetime. By contrast, a drug can be administered at any time and, in the brain, can be injected into specific areas.

The approach adopted by Gross *et al.*¹ is an ingenious way of addressing the shortcomings of gene knockouts, providing time- and tissue-specific deletion and restoration of serotonin_{1A} receptors. To achieve time-

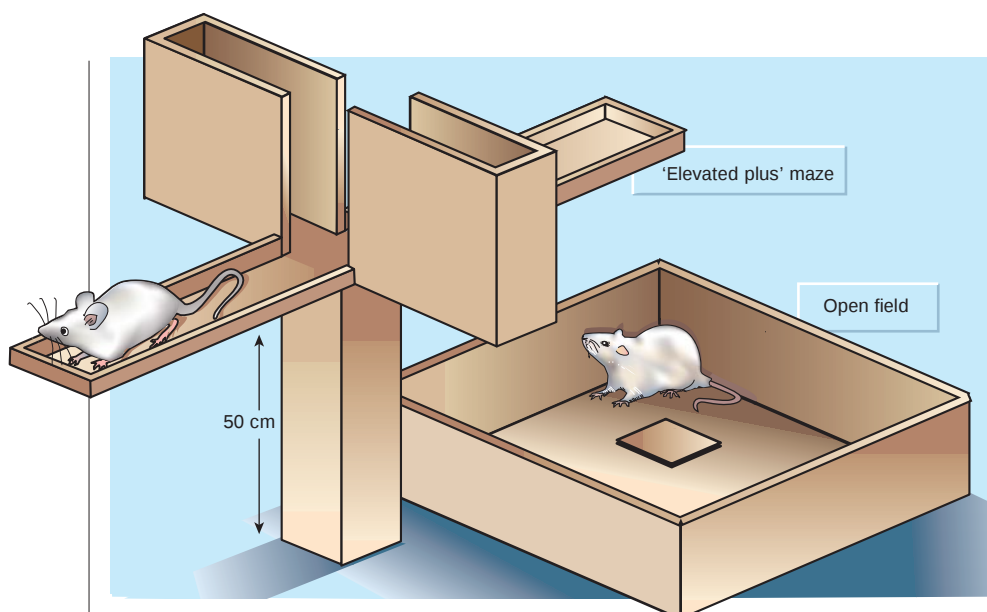


Figure 1 Over-anxious mice. Gross *et al.*¹ genetically modified mice so that the serotonin_{1A} receptor could be knocked out in a time- and tissue-specific manner. They then investigated the behaviour of the mice in several animal tests of anxiety, including the 'elevated plus' maze and the open field, shown here. In these tests, the mice face a conflict between their innate fear of an 'aversive' compartment (the open arms of the maze and the exposed centre of the open field) and their innate curiosity, which motivates them to explore the compartment. The authors monitored the mice for a variety of behaviours — such as how much time they spent exploring each compartment in the maze and the centre of the open field — some of which are taken to be a measure of 'anxiety'.

specific knockouts, Gross *et al.* produced mice in which expression of the serotonin_{1A}-receptor gene was under the control of the antibiotic doxycycline. The gene could be switched off — with a certain time lag — simply by feeding mice the antibiotic.

The authors also engineered mice in which the serotonin_{1A} receptor was expressed only in the forebrain. This was a little more complicated. They first produced a 'transgenic' mouse line in which the serotonin_{1A}-receptor gene could be activated only in the presence of a bacterial gene-transcription factor. In a second mouse line, the gene encoding that transcription factor was joined with a control region from the gene encoding calcium-calmodulin-dependent protein kinase II, which is active specifically in the forebrain. So the transcription factor is likewise expressed selectively in the forebrain. Mating the two lines of mice produced double-transgenic offspring with normal serotonin_{1A} receptors expressed in the hippocampus and cerebral cortex, but no receptors on the serotonin-sensitive neurons in the raphe nuclei of the hindbrain.

Gross *et al.* found that mice lacking serotonin_{1A} receptors throughout the brain showed pronounced anxiety-like behaviour, as judged in a variety of tasks (Fig. 1), whereas selective restoration of the receptors in the forebrain restored normal behaviour. Thus, serotonin_{1A} receptors in the forebrain regulate anxiety, whereas receptors in the hindbrain do not seem to be directly involved.

The authors also delineated critical periods of development during which the receptors influence anxiety. Feeding doxy-

cycline to juvenile (10–12-week-old) double-transgenic mice eliminated forebrain serotonin_{1A} receptors in adult mice. This did not elicit anxiety-like behaviour. By contrast, administering doxycycline during gestation eliminated the forebrain receptors from newborns, and this did lead to pronounced anxiety-like behaviour. So forebrain serotonin_{1A} receptors are needed during the development of newborns to modulate the predisposition to anxiety-like behaviour, but are no longer critical during adult life.

These findings¹, together with earlier evidence, suggest that selected serotonin-sensitive neuronal systems may participate in brain development. Like most neurotransmitter-sensitive neurons, serotonin-responsive neurons first appear during embryonic life and then gradually increase in numbers until adulthood. However, discrete but relatively prominent populations of such neurons are produced transiently at key periods during development and then vanish, suggesting that they may help to trigger particular aspects of brain development⁵. Moreover, the serotonin_{1A} receptor has been specifically implicated in the development of synapses (neuronal junctions) in the forebrain, as treatment of newborn rodents with drugs that block these receptors decreases the numbers of spines (neuronal extensions) on neurons in the dentate gyrus of the forebrain⁶.

More speculatively, perhaps variations in serotonin-sensitive neurons and serotonin receptors in early life account for the importance of maternal nurturing in preventing emotional disturbance in adults.



100 YEARS AGO

In a lecture lately delivered before the Norwegian Geographical Society by Captain R. Amundsen, the author gave an account of the proposed exploring expedition to the magnetic North Pole. Captain Amundsen was first officer of the *Belgica*, which sailed for the Antarctic in August 1897 with the view of determining the exact locality of the magnetic South Pole, and it was while that ship lay fixed in the drift-ice west of Graham Land that the idea was conceived of exploring the magnetic North Pole. For the contemplated expedition, the *Gjøa*... has been purchased at Tromsø. In 1831, Sir James Clark Ross reached a position where the dipping-needle was only deflected one minute from an absolutely vertical position, but the question has been raised whether the magnetic pole is actually only a point or whether the peculiarity of the needle assuming a vertical position extends over a large area, and, further, whether the magnetic pole changes its position. With the object of solving these two questions, Captain Amundsen will sail in the spring of 1903.

From *Nature* 27 March 1902.

50 YEARS AGO

In a previous note in these columns (*Nature*, 167, 260; 1951), we recorded the production by Messrs. Pest Control, Ltd., Harston, Cambridge, of a new systemic insecticide, *bis(isopropylamino)fluorophosphine oxide*, which was claimed to be safe for widespread use by the amateur gardener... It is now recommended that it be applied only in the form of capsules inserted in the soil and applied *only* to ornamental plants... It has been found that this same method of application, on a much larger scale, is highly successful in killing the mealybugs which transmit the virus of swollen shoot disease of cocoa in West Africa. The insecticide used in this case is termed 'Hanane', and it is now disclosed that this is, in fact, a mixture of the two well-known phosphorus insecticides produced by Schrader, *bis(dimethylamino)-fluorophosphine oxide* and *bis(bisdimethylamino) phosphonous anhydride*. The proposed plan is to apply this material to the roots of all those cocoa trees surrounding a focus of infection... So far there is no evidence available that these operations will, in fact, prevent the spread of the disease... But if the cocoa-growing industry of West Africa is to be saved, it is well worth while making an experiment.

From *Nature* 29 March 1952.

Thus, adult rats that were not licked and groomed adequately as pups by their mothers display increased anxiety, and this may reflect serotonin-dependent mechanisms⁷. Assuming that we can equate developmental stages in mice and humans, these findings might be relevant to brain development and the genesis of anxiety in people, too.

Serotonin appears to be an all-purpose neurotransmitter; it has been implicated in many aspects of brain function and in the effects of many important drugs that are used to treat anxiety, depression, migraine headaches, nausea, pain and irritable bowel syndrome. Gross *et al.*'s discovery¹ — that anxiety is linked to the need for serotonin_{1A} receptors in a specific brain region, at a particular period of development — adds a new layer of understanding of serotonin's function. More generally, the authors' technique lends greater precision and flexibility to

gene-knockout approaches for understanding neurotransmitter function, and will hopefully soon be extended to many other neurotransmitters and behaviours. ■

Solomon H. Snyder is in the Departments of Neuroscience, Pharmacology and Molecular Science and Psychiatry, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA.

e-mail: ssnyder@jhmi.edu

1. Gross, C. *et al.* *Nature* **416**, 396–400 (2002).
2. Parks, C. L., Robinson, P. S., Sibille, E., Shenk, T. & Toth, M. *Proc. Natl Acad. Sci. USA* **95**, 10734–10739 (1998).
3. Ramboz, S. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 14476–14481 (1998).
4. Heisler, L. K. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 15049–15054 (1998).
5. D'Amato, R. J. *et al.* *Proc. Natl Acad. Sci. USA* **84**, 4322–4326 (1987).
6. Heim, C. & Nemeroff, C. B. *Biol. Psychiatry* **49**, 1023–1039 (2001).
7. Yan, W., Wilson, C. C. & Haring, J. H. *Brain Res. Dev. Brain Res.* **98**, 185–190 (1997).

Astrobiology

Seeds of life?

Everett L. Shock

Amino acids, a basic constituent of life, can form in dust grains that are similar to those found in the space between stars. But how much does this tell us about the origins of life on Earth?

Organic compounds in meteorites and interplanetary dust particles are thought by some to hold the key to the origin of life. Increasingly, investigations are revealing a complex history of chemical

processes in the Solar System — processes by which the chemicals that are the basis of life may have been synthesized from the gas and dust that make up the interstellar medium. From page 401 of this issue, Bernstein *et al.*¹

and Muñoz Caro *et al.*² add another convolution to the plot: their experiments suggest that the conditions that are believed to exist on ice-coated grains in the interstellar medium may be suitable for the synthesis of amino acids, the building-blocks of proteins.

A wide range of conditions and reaction mechanisms can generate the spectrum of compounds observed in meteorites and dust particles. Isotopic evidence (especially enrichments in ¹⁵N and deuterium) supports the idea that at least half of the insoluble organic matter in carbonaceous meteorites originated from sources that existed before the Solar System formed^{3,4}. In contrast, soluble organic compounds in these meteorites, including amino acids, are generally thought to have been produced soon after the formation of the Solar System, by the action of aqueous fluids on the asteroids that gave rise to the meteorites (Fig. 1)⁵. In this scenario, the carbon, nitrogen and hydrogen that went into forming the amino acids found in the well-studied Murchison meteorite^{6–8} would have been carried by reactive precursors from the interstellar medium, but the amino acids themselves would have formed on the meteorite's parent body.

Results from the Tagish Lake meteorite⁹, which fell to Earth in 2000, make the picture more complicated still. Although it contains evidence of aqueous alteration, and although the bulk abundance of organic carbon is similar to that of the Murchison meteorite, concentrations of amino acids in the Tagish Lake meteorite are around a thousand times lower than those in Murchison.

Now two groups^{1,2} report amino-acid synthesis in experiments in which ice mixtures thought to be representative of the interstellar medium were subjected to ultraviolet radiation at temperatures below 15 K. Bernstein *et al.*¹ (page 401) report the synthesis of three amino acids (glycine, serine and alanine) in a mixture of water, methanol, ammonia and hydrogen cyanide, where the ratio of the components was 20:2:1:1, respectively. Muñoz Caro *et al.*² (page 403) report the synthesis of 16 amino acids and several other compounds in their 2:1:1:1:1 mixture of water, methanol, ammonia, carbon monoxide and carbon dioxide.

Assuming that conditions are comparable between the two experiments, the difference in their results demonstrates the profound effect that bulk composition can have on chemical synthesis. The water-rich experiment reported by Bernstein *et al.* produced only a few compounds. But the water-poor experiment of Muñoz Caro *et al.* generated a host of amino acids, some carrying not just one but two amino groups. The relative yields of mono-amino acids in this latter experiment show a reasonable correlation with the relative abundances found in extracts¹⁰ from the Murchison meteorite. But the meteorite shows no evidence of

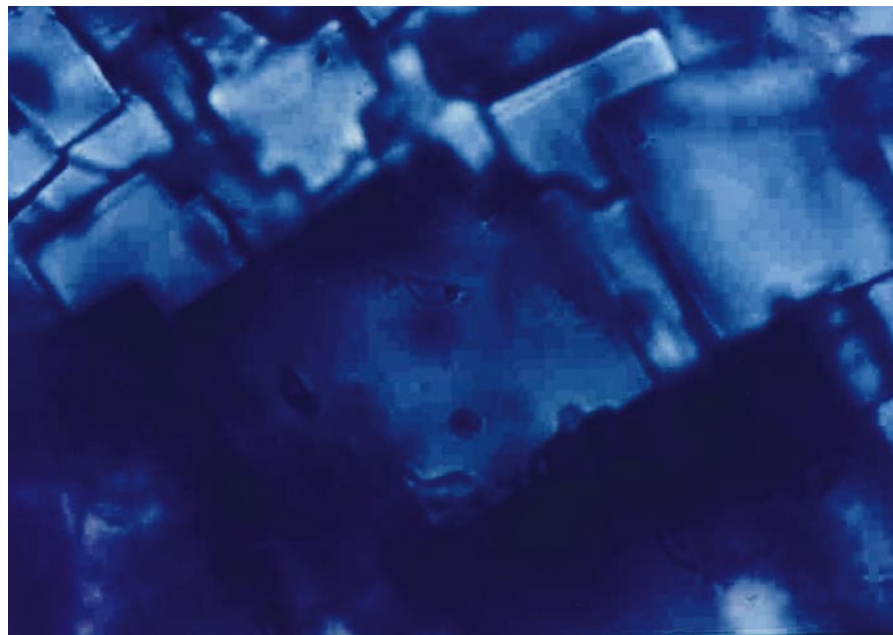


Figure 1 Is it possible that the amino acids found in meteorites originated in the interstellar medium, as recent laboratory experiments suggest? Or were they synthesized on the asteroid parent bodies of the meteorites by the action of aqueous fluids? This picture shows a fluid inclusion in salt crystals found in a meteorite that fell in Monahans, Texas, in 1998, and may be a rare sample of water from an asteroid.

di-amino acids such as those seen by Muñoz Caro and colleagues.

Of course, precisely matching meteorite compositions is not the goal of these experiments. Nor has it been the goal of dozens of other successful synthesis experiments, be they gas-phase reactions driven by spark discharges, ultraviolet radiation, shock waves, γ -rays or heat, or aqueous synthesis from a variety of starting materials, including hydrogen cyanide, formaldehyde, methane, carbon monoxide, oxalic acid, carbides or mixtures of these compounds. It appears that nearly every experimental scenario produces organic compounds of some form, which could hardly be more frustrating when trying to unravel what actually produced the organic inventory of the Solar System. Perhaps it is time to stop this seemingly endless series of phenomenological experiments and to concentrate instead on quantifying reaction rates, the relative stabilities of synthesized compounds and the effects of variables such as temperature and pressure.

What about the big picture? Given the truly diverse array of energy sources that may initiate synthesis, complex organic compounds should be expected in any sector of the Solar System that is rich in volatile elements. This is especially true for the outer parts of the Solar System, which host most of the light elements that did not go into making the Sun. On Earth, the hydrosphere (water, including ice and water vapour) accounts for 0.0013% of the planetary volume. But water makes up more than 15% of the volume of Europa, and more than 65% of that of Ganymede — two of Jupiter's largest moons. Even the drier carbonaceous meteorites show evidence of a water-rich past on their parent bodies; some have organic carbon contents that rival those of petroleum source rocks on Earth. So it is plausible to suggest that most organic compounds in our Solar System are extraterrestrial — and abiological — in origin.

Does this tell us much about the origin of life? Well, you can study geology for a living, but knowing how different rocks form doesn't tell you which lumps of rock will become Teotihuacán, the Taj Mahal or Tony's Tavern. Studying the chemical building-blocks of life shows that they are ubiquitous and can exist in the absence of life. Indeed, inferred cosmic abundances of these building-blocks from abiological sources greatly exceed those from living organisms. Accepting that fact, it follows that process-driven investigations into the emergence of life may need to be cast in a different way, which takes into account the materials involved but is not directly tied to them. This, I believe, is a major challenge for the fledgling field of astrobiology. ■

Everett L. Shock is in the Department of Earth and Planetary Science, Washington University, St Louis, Missouri 63130, USA.

e-mail: shock@zonvark.wustl.edu

- Bernstein, M. P., Dworkin, J. P., Sandford, S. A., Cooper, G. W. & Allamandola, L. J. *Nature* **416**, 401–403 (2002).
- Muñoz Caro, G. M. *et al. Nature* **416**, 403–406 (2002).
- Messenger, S. *Nature* **404**, 968–971 (2001).
- Alexander, C. M. O'D. *et al. Meteor. Planet. Sci.* **33**, 603–622 (1998).
- Cronin, J. R. & Chang, S. in *The Chemistry of Life's Origins* (eds Greenberg, J. M., Mendoza-Gomez, C. X. & Pirronello, V.)

- 209–258 (Kluwer, Dordrecht, 1993).
- Engel, M. H., Macko, S. A. & Silfer, J. A. *Nature* **348**, 47–49 (1990).
- Pizzarello, S., Krishnamurthy, R. V., Epstein, S. & Cronin, J. R. *Geochim. Cosmochim. Acta* **55**, 905–910 (1991).
- Engel, M. H. & Macko, S. A. *Nature* **389**, 265–268 (1997).
- Pizzarello, S. *et al. Science* **293**, 2236–2239 (2001).
- Shock, E. L. & Schulte, M. D. *Geochim. Cosmochim. Acta* **54**, 3159–3173 (1990).

Cell biology

The ubiquitin connection

Howard Riezman

The cellular world has some daunting problems for biologists. For example, cells use the ubiquitin molecule in different ways to achieve different effects. These tangled events come under the spotlight in new work.

The cells in our bodies need to interact with each other and with the environment around them. For example, they take up certain nutrients (such as cholesterol, vitamin B₁₂ and iron) from their environment, and can signal to each other with messenger molecules that activate receptor proteins on the cell surface. Endocytosis — the process by which cells invaginate and then internalize portions of their own outer membrane — is central to these, and many other, cellular events. It allows cells to engulf and then take up nutrients, and to regulate cell-to-cell signalling by internalizing activated receptors.

Several proteins are involved in controlling endocytosis, and some of these are distinguished by the fact that they become labelled with a single copy of the ubiquitin molecule. On page 451 of this issue, Polo and colleagues¹ look at how this labelling comes about. Their results shed some light on the broader question of how cells differentiate between proteins that are to be tagged with just one or many copies of ubiquitin —

events with very different outcomes. They also suggest how certain receptor proteins may be 'sorted' into invaginations, ready to be internalized.

The internalization step of endocytosis has been studied intensively, and the results have contributed greatly to our understanding of membrane-transport processes in general. An intracellular protein complex called clathrin is proposed to drive the formation of membrane invaginations ('pits') by coating portions of the membrane. Clathrin-coated pits are then pinched off to form clathrin-coated transport vesicles inside the cell². To get into pits, many cell-surface proteins (such as receptors) in animal cells carry short peptide signals on their intracellular tails that interact directly with components of the clathrin coat.

But this is not universal. For example, many yeast plasma-membrane proteins are covalently modified on specific lysine amino acids by one (or at most two) ubiquitin molecules, and this seems to be the signal driving the proteins towards endocytosis. (Note that

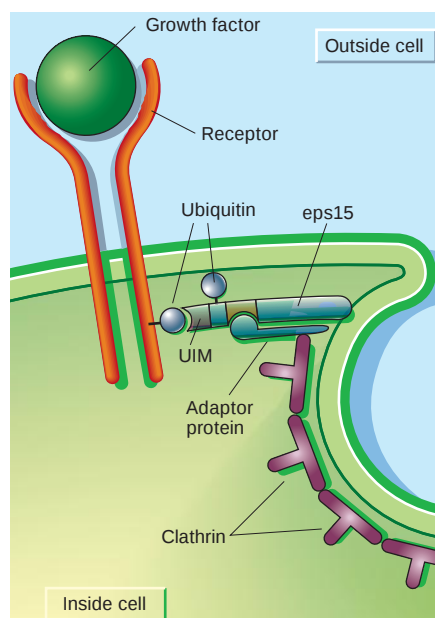


Figure 1 The many uses of ubiquitin in endocytosis. Polo *et al.* looked at how various proteins that regulate endocytosis, including eps15, become modified with a single ubiquitin protein. They found that these proteins have a specific region that is necessary for such monoubiquitination, and that the previously identified ubiquitin-interacting motif (UIM) forms part of this region and is essential for monoubiquitination. They propose that this region, by binding to other ubiquitinated proteins such as activated growth-factor receptors, could recruit those proteins into membrane pits that are coated with the clathrin protein complex (these pits will later pinch off to form intracellular transport vesicles). Adaptor proteins link eps15 to clathrin.

Daedalus

Learning in middle life

The brain can learn; indeed, that is its main purpose. But with a mere ten billion cells it must use space with great economy. A whole lifetime of learning, language and experience must be packed in, all with almost instant recall. By contrast, a computer can easily store ten billion bits, but wastes nearly all of them.

This reminds Daedalus of last week's scheme to erase chosen bits of the brain. Botulinum toxin was injected into areas of skin, chosen by psychiatric acupuncturists to home in on selected brain areas. The toxin was encapsulated in small particles to travel slowly up the local nerve and reach the relevant brain area as the encapsulation decayed. The obsession, troublesome memory, complex or whatever, would be erased by the toxin for several months. One snag is that people with many obsessions and problems could lose data. The vacant cells would resemble a piece of empty brain tissue. Daedalus now reckons that, as it becomes available, the brain will seize this extra capacity avidly, and fill it with new data. The erased cells will more than recover. Even unlearning, the hardest feat of all, will become possible.

Unlearning and relearning are important these days. Increasingly, we move to new jobs and new partners; 'lifetime education' is widely advocated; several languages are now important. Yet we only have one small brain. How valuable to erase sizeable chunks of information, and fill the empty cells with new data! Our early formal education, for example, now irrelevant and largely overlaid with newer understanding, must still be in there somewhere, if only we could get it out.

So DREADCO is developing mid-life brain erasure. It could become as common as cosmetic surgery. Daedalus's volunteers will discover the most fruitful regions for forgetting, and the areas of skin that access them, by botulinum treatment. Such major innovations as the euro require huge acts of forgetting. New computer techniques are famously the province of the young, who do not have to unlearn the old stuff. Some businesses are so new that experience is a positive disadvantage. But now even the oldsters, with a lifetime's experience packed into their brains, could benefit. The selective amnesia of the botulinum treatment would enable them to learn new facts and skills. The dire verdict on so many politicians, "Where there's death there's hope", might not be so final.

David Jones

these proteins are not the same as those ubiquitinated proteins that actually regulate endocytosis.) Mammalian cells probably use ubiquitination, too³. Yet no obvious connections between ubiquitin and clathrin have been found, so it has been unclear how ubiquitinated membrane proteins are sorted into clathrin-coated pits.

This is one of the questions to which Polo *et al.*¹ provide a possible answer. But this is really an added extra — the main focus of their work was what determines how endocytosis-regulating proteins become labelled with just one rather than many ubiquitin molecules ('monoubiquitination' versus 'polyubiquitination'). This is a crucial distinction, because proteins that are tagged with longer ubiquitin chains (at least four ubiquitins) are destined for destruction: they are recognized by the proteasome, a large, barrel-shaped protein complex that destroys cytoplasmic proteins. Yet little is known about how this distinction is made. Polo *et al.* approached the problem by studying the mechanism by which proteins that regulate endocytosis, including eps15 and eps15R, are monoubiquitinated.

The authors started by mapping regions in eps15 and eps15R that are required for their monoubiquitination. They found that these regions overlap with a previously known sequence of amino acids called the ubiquitin-interacting motif (UIM)⁴. By mutating single amino acids within this motif, Polo *et al.* found that it is required both to allow these proteins to interact with ubiquitin, and for their monoubiquitination (although the motif is not itself ubiquitinated).

Ubiquitination is a complex process and usually requires three sequentially acting enzymes. The last enzyme in this series is a ubiquitin ligase, which transfers ubiquitin to the target protein. Polo *et al.* identified such an enzyme, Nedd4, that can add a single ubiquitin molecule to eps15. This enzyme has already been implicated in the ubiquitination of plasma-membrane proteins, including a sodium channel⁵, that are downregulated by endocytosis. But it seems that Nedd4, like its yeast relative Rsp5, might be involved in both monoubiquitination and polyubiquitination⁶. How does it 'decide' what to do? One possibility is that other proteins regulate Nedd4 and related ligases. Another possibility — one of two discussed by Polo *et al.*¹ — is that once a protein with a UIM is monoubiquitinated, the UIM binds to the ubiquitin group and prevents further modification. This mechanism may not apply to all monoubiquitinated proteins; they do not all have a UIM. Nevertheless, it is possible that a UIM on one protein could act in 'trans' by binding to the ubiquitin on another monoubiquitinated protein.

Interestingly, the eps15 protein can interact with adaptor proteins that are attached to clathrin-coated membrane pits. The fact

that it contains a UIM suggests that eps15 might also interact with other ubiquitinated plasma-membrane proteins, including receptors, that cannot themselves interact with clathrin, pulling those proteins into pits (Fig. 1). So eps15 might be one of the sought-after connections between ubiquitin molecules, used as endocytic signals on plasma-membrane proteins, and clathrin.

Eps15 and eps15R are not the only proteins involved in endocytosis — whether in internalization or in other steps — that contain a UIM. Two other examples are the epsin and Hrs proteins, which are also monoubiquitinated, as Polo *et al.*¹ show. Epsins interact directly with clathrin⁷ and are involved in internalization. Hrs is involved in recruiting clathrin to endosomes⁸, cellular compartments to which internalized material is targeted before moving further along the endocytic pathway. Epsins and Hrs share another characteristic: they both have domains that bind to specific lipids. Epsins have an 'ENTH' domain, which can bind to phosphatidylinositol-4,5-bisphosphate⁹, a lipid that can be generated at the plasma membrane. Hrs has a 'FYVE' domain that can bind to phosphatidylinositol-3-phosphate, which is generated on endosomes¹⁰. The use of many types of interaction — between different proteins, between proteins and lipids, and between proteins and ubiquitin — may be necessary to determine exactly where clathrin-coated vesicles form on cell membranes.

There is much still to be learned about monoubiquitination and UIMs. For example, these motifs might be involved in regulating the monoubiquitination of other proteins, such as cell-surface receptors. One can also speculate that monoubiquitination of a target UIM-containing protein leads that protein to release a ubiquitin-containing partner, freeing it to bind another protein. Monoubiquitination may also work like other protein modifications, such as phosphorylation, to induce a conformational change in the target protein, regulating its activity directly. And UIMs could allow the construction of larger protein networks, particularly as many proteins have more than one such motif. All of this remains to be seen: Polo *et al.*'s work¹ marks the beginning of an exciting quest.

Howard Riezman is at the Biozentrum of the University of Basel, CH-4056 Basel, Switzerland.
e-mail: howard.riezman@unibas.ch

1. Polo, S. *et al.* *Nature* **416**, 451–455 (2002).
2. D'Hondt, K., Heese-Peck, A. & Riezman, H. *Annu. Rev. Genet.* **34**, 255–295 (2000).
3. Hicke, L. *Nature Rev. Mol. Cell Biol.* **2**, 195–201 (2001).
4. Hofmann, K. & Falquet, L. *Trends Biochem. Sci.* **26**, 347–350 (2001).
5. Rotin, D., Staub, O. & Hagenauer-Tsapis, R. *J. Membr. Biol.* **176**, 1–17 (2000).
6. Hicke, L. *Cell* **106**, 527–530 (2001).
7. Rosenthal, J. A. *et al.* *J. Biol. Chem.* **274**, 33959–33965 (1999).
8. Raiborg, C. *et al.* *EMBO J.* **20**, 5008–5021 (2001).
9. Itoh, T. *et al.* *Science* **291**, 1047–1051 (2001).
10. Simonsen, A., Wurmser, A. E., Emr, S. D. & Stenmark, H. *Curr. Opin. Cell Biol.* **13**, 485–492 (2001).

Obituary

Sir Raymond Firth (1901–2002)

With the death of Raymond Firth on 22 February, anthropology has lost one of its giants. Firth's 80-year career encompassed the development of the modern form of the subject. He left his mark on almost every sub-field in his speciality, social anthropology.

Firth was born in 1901 in Auckland, New Zealand, and was educated in economics at Auckland University College. His interest in anthropology began when, as a schoolboy browsing through an Auckland bookshop, he came upon Frederic Maning's *Old New Zealand*, a study of traditional Maori life. So began Firth's lifelong fascination with the native peoples of the South Pacific, an interest that would eventually inspire his first book, *Primitive Economics of the New Zealand Maori*, published in 1929.

In the mid-1920s, Firth moved to London to pursue a doctorate at the London School of Economics, where he came under the influence of Bronislaw Malinowski: Malinowski's meticulous studies of the Trobriand islanders of New Guinea were among the earliest fieldwork-based ethnographies. This was a time when the basic concepts and approaches of anthropology were being formulated. French sociology was under the sway of Émile Durkheim's concern with 'social solidarity', while the philosopher Lucien Lévy-Bruhl was investigating cultural effects on thought and logic. Across the Atlantic, Franz Boas and Alfred Kroeber emphasized the concept of 'culture', a collective system of beliefs and practices. Over the next several decades, Boas's students Margaret Mead and Ruth Benedict published best-sellers portraying cultures boldly as contrasting psychological types.

British social anthropologists, by contrast, were not interested in broad speculations about mind or society. Nor were they concerned with patterns of culture. Instead, they produced fine-grained accounts of a society's functional integration, derived from empirical studies of small-scale societies. Their general concern was with social structure, rather than culture or mind.

In 1928, Firth returned to the Pacific for a year's fieldwork on the tiny island of Tikopia at the eastern end of the Solomon Islands. Although, geographically, Tikopia lies within Melanesia, it is a Polynesian outlier, at that time home to some 1,300 Polynesians. This trip inaugurated Firth's long relationship with Tikopia and its people, the eventual result being dozens of articles and nine books — the greatest



A Tikopian odyssey

record of any single society by a single ethnographer. In 'simple' societies with relatively little institutional specialization, ethnography was inevitably a generalist enterprise, demonstrating how specific aspects of social life were interlocking pieces of a larger puzzle. Firth would later call the big picture 'social organization' and would stress its dynamics.

We, *The Tikopia* (1936), Firth's first book on Tikopia, is still a classic, notable for its mix of detailed observation and gentle theoretical insight. Firth made little Tikopia into a 'big place' for anthropology, demonstrating the value of paying a lifetime of attention to a small, isolated and coherent community. Firth's last book on Tikopia, a study of the island's music, was published in 1990. In between, there were volumes on religion, politics and economics, equivalent to several pages for every living Tikopian.

Malinowski had championed a kind of 'universalist functionalism' that asked how basic human needs were met in different social settings. Firth preferred a more relativistic vision of social function, stressing how human needs might be locally shaped. For example, following a long discussion of the prohibition of incest in *We, The Tikopia*, Firth concluded "that the incest situation varies according to the social structure of each community, that it has little to do with the prevention of sex relations as such, but that its real

correlation is to be found in the maintenance of institutional forms in the society as a whole, and of the specific interest of groups in particular". But although emphasizing social institutions, Firth was also sensitive to the importance of individuals, to the role of personal choice and to the flexibility of social norms.

Firth also published influential studies of Malay fishermen and middle-class kinship in London, and he co-edited a multi-volume handbook on the South Pacific for British naval intelligence during the Second World War. He also wrote several general works, notably *Human Types: An Introduction to Social Anthropology* (1938), *Elements of Social Organisation* (1951) and, in 1996, *Religion: A Humanist Interpretation*.

Apart from a two-year spell at the University of Sydney in the early 1930s, Firth's academic home remained the London School of Economics. He was appointed professor in 1944 and remained there until his retirement in 1968.

In the following years he became a peripatetic teacher, undertaking lecturing stints at several universities. In 1971–72 he was visiting professor at the University of Chicago, where 'symbolic anthropology' was being developed. Unfazed by fears that he might be bringing coals to Newcastle, Firth delivered a series of engaging lectures on human symbolism which were published in 1973 as *Symbols: Public and Private*.

It was at Chicago that I first encountered Raymond Firth in person. As a student in his 'Post-Field Seminar', I saw a master teacher at work. He was an intent listener with a keen memory for the contributions of each participant. Firth choreographed his seminars, calling upon particular students to develop a thread of discussion based on what they had said previously. Inevitably, a coherent and engaging debate would emerge. Firth's teaching skills were clearly tied to his formidable abilities as an ethnographer.

Raymond Firth will be remembered for many virtues: boundless energy, an infectious curiosity, a commitment to getting the facts right, a talent for finding big things in small places, and his sheer durability as a scholar. Temperamentally, Firth was notable for his intellectual balance and his unflagging good sense. Personally, he was a thoroughly delightful and genial gentleman, with a talent for long friendships.

His wife of 66 years, Rosemary, died last year; she, too, was an anthropologist and a frequent collaborator with her husband. A son, Hugh, survives them.

Bradd Shore
Bradd Shore is in the Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA.
e-mail: antbs@emory.edu

Spinning eggs — a paradox resolved

An explanation for an odd egg performance is rolled out in time for Easter.

If a hard-boiled egg is spun sufficiently rapidly on a table with its axis of symmetry horizontal, this axis will rise from the horizontal to the vertical. (A raw egg, by contrast, when similarly spun, will not rise.) Conversely, if an oblate spheroid is spun sufficiently rapidly with its axis of symmetry vertical, it will rise and spin about the vertical on its rounded edge with its axis of symmetry now rotating in a horizontal plane. In both cases, the centre of gravity rises; here we provide an explanation for this paradoxical behaviour, through derivation of a first-order differential equation for the inclination of the axis of symmetry.

The behaviour of these spheroids shows similarities to that of the mushroom-shaped top (known as the *tippe-top*) that has been previously analysed^{1–3}; however, there are also significant differences that arise from the geometry of the general axisymmetric body. The essential mechanism can be traced to the action of the frictional force between the spinning object and the table. If this frictional force vanishes, or if the body rolls without slipping, the ‘rise’ phenomenon does not occur. Weak friction associated with slipping at the point of contact causes a secular instability, which gives rise to the surprising behaviour that is observed. (A fuller discussion of the subtleties involved is in preparation.)

The geometry and notation for the problem are shown in Fig. 1. We seek to obtain an equation for the polar angle, $\theta(t)$, between Oz and OZ . We adopt a rotating frame of reference $OXYZ$, with OX horizontal in the plane Π and OY into the plane of the diagram. The angular velocity of the body, ω , has components in this frame

$$\omega = ((n - \Omega \cos \theta) \sin \theta, \dot{\theta}, \Omega \sin^2 \theta + n \cos \theta) \quad (1)$$

where $n(t)$ is the spin (that is, the component of ω about Oz), and the dot represents time-differentiation. The angular momentum, $\mathbf{H}$, has components

$$\mathbf{H} = ((Cn - A\Omega \cos \theta) \sin \theta, A\dot{\theta}, A\Omega \sin^2 \theta + Cn \cos \theta) \quad (2)$$

where A , A and C are the principal moments of inertia at O . The height, $h(\theta)$, of O above the table is determined as a function of θ from the geometry and density distribution of the body. The components of the vector $\mathbf{X}_p = (X_p, 0, Z_p)$ from O to the point of contact, P , are then given by

$$X_p = dh/d\theta, Z_p = -h(\theta) \quad (3a, b)$$

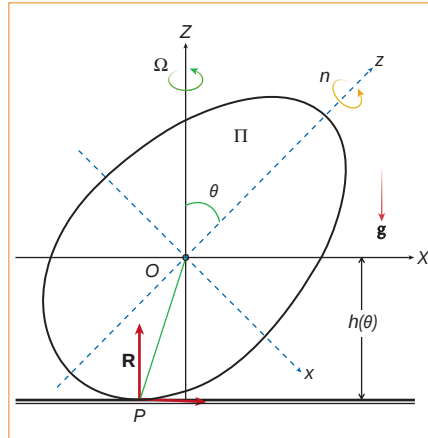


Figure 1 An axisymmetric body with centre of mass O spins on a horizontal table with point of contact P . Its axis of symmetry, Oz , and the vertical axis, OZ , define a plane Π (containing OP), which precesses about OZ with angular velocity $\Omega(t) = (0, 0, \Omega)$. $OXYZ$ is a rotating frame of reference, with OX horizontal in the plane Π . The height of O above the table is $h(\theta)$ and the coordinates of P are $X_p = dh/d\theta$, $Y_p = 0$, $Z_p = -h(\theta)$. The forces acting on the body are its weight $M\mathbf{g} = (0, 0, -Mg)$, the normal reaction $\mathbf{R}$ at P , and the frictional force $\mathbf{F}$ at P in the Y -direction.

If the frictional force vanishes, the horizontal velocity components of O may be taken to be zero (as no horizontal force then acts on the body). The slip velocity of the point P is then $\mathbf{U}_p = (-h\dot{\theta}, V_p, 0)$, where

$$V_p = (\Omega \sin^2 \theta + n \cos \theta) dh/d\theta + (n - \Omega \cos \theta) h(\theta) \sin \theta \quad (4)$$

In any steady state, $\mathbf{U}_p = 0$. When the frictional force is weak, $\dot{\theta}$ is correspondingly small (as may be verified retrospectively), and at leading order, $\mathbf{U}_p = (0, V_p, 0)$. The frictional force, $\mathbf{F}$, tends to resist slipping; it follows that at leading order, $\mathbf{F} = (0, F, 0)$, where F is a function of V_p given by the law of dynamic friction between the two surfaces in contact. There is also a normal reaction $\mathbf{R} = (0, 0, R)$ at P , where R is of order Mg .

Euler's angular-momentum equation is

$$\partial \mathbf{H} / \partial t + \Omega \times \mathbf{H} = \mathbf{X}_p \times (\mathbf{R} + \mathbf{F}) \quad (5)$$

the right-hand side representing the torque exerted on the body relative to O . The Y -component of this equation is

$$A\ddot{\theta} - A\Omega^2 \cos \theta \sin \theta + Cn\Omega \sin \theta = -RX_p \quad (6)$$

Now $|\ddot{\theta}| \ll \Omega^2$, because the secular change of θ with which we are concerned is slow; hence the first term of equation (6) may be neglected. Moreover, we shall suppose, for simplicity, that Ω^2 is sufficiently large that the terms that involve Ω in equation (6) dominate the term $-RX_p$. The precise condition for the case of uniform spheroids is given below (condition (14)). Thus, at leading order, equation (6) becomes simply $(Cn - A\Omega \cos \theta)\Omega \sin \theta = 0$, and so, provided that $\sin \theta \neq 0$,

$$Cn = A\Omega \cos \theta \quad (7)$$

a condition that we may describe by the term ‘gyroscopic balance.’ In adopting this approximation (which is analogous to the geostrophic approximation in fluid mechanics), oscillatory modes of high frequency ($O(\Omega)$) are filtered from the system. Using equation (7), the angular momentum (equation (2)) is simplified to $\mathbf{H} = (0, A\dot{\theta}, A\Omega)$.

The Z - and X -components of equation (5) now give, respectively,

$$A\dot{\Omega} = FX_p, A\Omega\dot{\theta} = FZ_p \quad (8a, b)$$

from which we can immediately obtain

$$\dot{\Omega}/\Omega = (X_p/Z_p)\dot{\theta} = -\dot{h}/h \quad (9)$$

using equation (3). This integrates to give $\Omega h = \text{cst.}$, or equivalently,

$$-\mathbf{H} \cdot \mathbf{X}_p = A\Omega h = J \quad (10)$$

where J is a constant determined by the initial conditions. This ‘Jellet’ constant has been found previously⁴ for the particular case of bodies (such as the *tippe-top*) on the assumption that the portion of the surface in contact with the plane is spherical. The above derivation shows that the constant still exists quite generally as an ‘adiabatic invariant’ under the gyroscopic-balance condition (7) for the case of the general axisymmetric body. From equations (4), (7) and (10), V_p is now known as a function of θ ; this may be expressed compactly as

$$V_p = (J/A)\zeta^{-3}h^{-1}d(\zeta h)/d\theta \quad (11)$$

where $\zeta(\theta) = (\sin^2 \theta + (A/C)\cos^2 \theta)^{-1/2}$.

Substituting equations (3b) and (10) into equation (8b) now gives

$$J\dot{\theta} = -Fh^2(\theta) \quad (12)$$

If the frictional relationship between F and V_p is known, and if $h(\theta)$ is known through geometrical considerations, then equation (12) may be integrated to determine θ as a

function of t . A steady state can be attained only when $V_p = 0$ (and so $F = 0$), because only then does the energy dissipation due to friction at P vanish.

For example, suppose that the body is the uniform spheroid $a^2(x^2 + y^2) + b^2z^2 = a^2b^2$. The principal moments of inertia at its centre, O , are $A = M(a^2 + b^2)/5$ and $C = 2Mb^2/5$. The function $h(\theta)$ is easily determined in the form $h(\theta) = (a^2\cos^2\theta + b^2\sin^2\theta)^{1/2}$, and from equation (11) we obtain the remarkable simplification

$$V_p = -(J/4Ah^2)(a^2 - b^2)\sin 2\theta \quad (13)$$

Gyroscopic balance occurs provided that

$$\Omega^2 \gg \frac{5g|a^2 - b^2|}{(a^2 + b^2)\min(a, b)} \quad (14)$$

As regards the frictional force, there are two obvious possibilities. First, we may assume a Coulomb law, $F = -\mu MgV_p/|V_p|$; then equation (12) integrates to give

$$\tan \theta = -(a/b)\tan \mu q(t - t_0) \quad (15)$$

where $q = Mgab(a - b)/|a - b|J$ and t_0 is a constant of integration. Here we may take $J = A\Omega^*/((a^2 + b^2)/2)^{1/2}$, where Ω^* is the value of Ω when $\theta = \pi/4$. For the prolate spheroid ($q > 0$), this describes a monotonic decrease of θ from $\pi/2$ to 0 over the time interval $[t_0 - \pi/2\mu q, t_0]$; for the oblate spheroid ($q < 0$), it similarly describes a monotonic increase of θ from 0 to $\pi/2$ over the time interval $[t_0, t_0 + \pi/2|\mu q|]$. In either case, equation (15) thus represents a transition from the unstable to the stable state over a finite time interval $\Delta t = \pi/2|\mu q|$.

Alternatively, if we assume a 'viscous'

friction law $F = -\mu MgV_p$, then equation (12) becomes $\dot{\theta} = -(\mu q/2)\sin 2\theta$, where now $q = 5g(a^2 - b^2)/2(a^2 + b^2)$. This integrates to give

$$\tan \theta = e^{-\mu q(t - t_0)} \quad (16)$$

which again describes monotonic transition from unstable to stable states (whether $q > 0$ or $q < 0$) on the secular time scale $|\mu q|^{-1}$. Numerical integration of the exact equations (without use of the gyroscopic approximation) confirms the validity of this description.

The key to the above analysis lies in the fact that the Jellett constant (equation (10)) exists for arbitrary bodies of revolution as an adiabatic invariant when the condition described by equation (7) is satisfied, and not only for the previously analysed bodies with part-spherical forms.

Finally, we may note that a raw egg does not rise when spun, simply because the angular velocity imparted to the shell must diffuse into the fluid interior; this process dissipates most of the initial kinetic energy imparted to the egg, the remaining energy being insufficient for condition (14) to be satisfied and for the state of gyroscopic balance to be established.

H. K. Moffatt*, Y. Shimomura†

*Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, UK

e-mail: hkm2@damtp.cam.ac.uk

†Department of Physics, Keio University, Hiyoshi, Yokohama 223-8521, Japan

e-mail: yutaka@phys-h.keio.ac.jp

1. Braams, C. M. *Physica* **18**, 503–514 (1952).
2. Hugenholtz, N. M. *Physica* **18**, 515–527 (1952).
3. Gray, C. G. & Nickel, B. G. *Am. J. Phys.* **68**, 821–828 (2000).
4. Jellett, J. H. *A Treatise on the Theory of Friction* (Macmillan, London, 1872).

Competing financial interests: declared none.

Evolutionary biology

Lamprey *Hox* genes and the origin of jaws

The development of jaws was a critical event in vertebrate evolution because it ushered in a transition to a predatory lifestyle, but how this innovation came about has been a mystery. In the embryos of jawed vertebrates (gnathostomes), the jaw cartilage develops from the mandibular arch, where none of the *Hox* genes is expressed; if these are expressed ectopically, however, jaw development is inhibited^{1–3}. Here I show that in the lamprey, a primitively jawless (agnathan) fish that is a sister group to the gnathostomes⁴, a *Hox* gene is expressed in the mandibular arch of developing embryos. This finding, together with outgroup comparisons, suggests that loss of

Hox expression from the mandibular arch of gnathostomes may have facilitated the evolution of jaws.

In gnathostome embryos, cranial neural-crest cells migrate from the overlying mid- and hindbrain into the pharyngeal arches^{5,6}, where they give rise to the pharyngeal skeleton, including the jaw. With the exception of those destined for the mandibular arch, cranial neural-crest cells express specific combinations of *Hox* genes, which reflect their site of origin in the segmented hindbrain^{5,6}. In lamprey embryos, the branchial skeleton also develops from cranial neural-crest cells^{7,8} but, in contrast to gnathostomes, the lamprey mandibular arch fails to form separate dorsal and ventral cartilage condensations^{4,9}. Instead of forming a Meckel's cartilage, the lamprey first arch forms the velum, a muscular pumping organ

with a small, solitary cartilaginous condensation⁴ (Fig. 1a).

To investigate whether modulation of *Hox*-gene expression in the mandibular arch could be associated with the evolutionary origin of jaws, I cloned lamprey *Hox* genes and analysed their expression patterns during embryogenesis. Two overlapping genomic cosmid (MPMGc05511883 and MPMGc055J22138) were found to contain three *Hox* genes with orthology to the *Hox5-6-7* cognate group. During lamprey embryogenesis, the *HoxL6* (GenBank accession no. AY089982) expression domain spreads anteriorly from the blastopore and extends into the cranial mesoderm (see supplementary information), where I detected transcripts from the onset of mandibular-arch formation prior to colonization of this region by neural-crest cells⁹ (Fig. 1b).

HoxL6 continues to be expressed strongly in the mandibular and posterior arches as these become individually defined (Fig. 1c), and in two stripes in the dorsal hindbrain (see supplementary information), where the neural crest originates^{7–9}. In older embryos, *HoxL6* expression was detected within each pharyngeal arch, and in the upper and lower lips (Fig. 1d). *HoxL6* transcripts co-localize with *Dlx*, a marker of the lamprey cranial neural crest^{10,11}, and were later detected in hypertrophic chondrocytes, indicating that *HoxL6* is expressed by neural-crest cells in the pharyngeal arches (compare Fig. 1d and e, and see supplementary information).

Within the mandibular arch, however, I detected *HoxL6* transcripts and *Dlx* protein in mutually exclusive ventral and dorsal domains, respectively (Fig. 1d, e). Comparison of *HoxL6* and *HoxL5* (GenBank accession no. AY089981) expression revealed that the *HoxL6* domain extends anterior to the boundary of *HoxL5* expression (see supplementary information), suggesting that spatial collinearity¹² has been broken. These findings identify lampreys as the only known vertebrate in which *Hox* genes are expressed in the mandibular arch.

To determine whether the breakage of *Hox6* collinearity is a derived condition of lampreys, I analysed expression of the single *Hox6* orthologue *AmphiHox6* in the cephalochordate amphioxus. *AmphiHox6* expression was detected throughout the neural tube, up to the base of the cerebral vesicle, and in endoderm up to the first gill slit (Fig. 1f). *AmphiHox6* expression extends anterior to the *AmphiHox3* and *AmphiHox4* domains¹² and, like *AmphiHox2*, does not respect spatial collinearity. Thus, in cephalochordate and lamprey embryos, *Hox6* genes are expressed anterior to their 3' neighbours, indicating that this could be a general feature of non-

Prey attack by a large theropod dinosaur

Prey-capture strategies in carnivorous dinosaurs have been inferred from the biomechanical features of their tooth structure, the estimated bite force produced, and their diet^{1–3}. Rayfield *et al.*⁴ have used finite-element analysis (FEA) to investigate such structure–function relationships in *Allosaurus fragilis*, and have found that the skull was designed to bear more stress than could be generated by simple biting. They conclude that this large theropod dinosaur delivered a chop-and-slash ‘hatchet’ blow to its prey, which it approached with its mouth wide open before driving its upper tooth row downwards. We argue that this mode of predation is unlikely, and that the FEA results, which relate to an ‘overengineered’ skull, are better explained by the biomechanical demands of prey capture. Understanding the mechanics of predation is important to our knowledge of the feeding habits of carnivorous dinosaurs and for accurate reconstruction their lifestyles.

First, we note that no living carnivorous tetrapod attacks prey in quite the way that the authors contend. For example, in lizards that grasp their prey without the aid of the tongue, the attack depends upon careful and precise jaw closure, which may be followed by a shaking or lateral flailing of the prey; it does not depend on an initial high-impact collision^{5,6}.

Second, if *Allosaurus* is unusual in using such a biomechanically stressful chop/slash attack, then its tooth morphology, and possibly its tooth regionalization, should reflect this, and its jaw and tooth design should be substantially different from that of other carnosaurs. This is not the case, however, as allosaur teeth are unremarkable among theropod dinosaurs^{2,7}. In fact, the teeth of the upper jaw (which would deliver the blow) are not much different from those of the lower jaw (which would not). The modest tooth regionalization in the allosaur specimen cited by Rayfield *et al.*⁴ is comparable to that of other carnosaurs, including *Tyrannosaurus*, *Albertosaurus* and *Velociraptor*, that presumably do not use a hatchet-like attack.

Third, and perhaps most revealing, the skull of *Allosaurus* was kinetic — equipped with a movable basal joint⁷ (Fig. 1). Kinetic skulls occur widely among non-mammalian tetrapods, including the earliest⁸. In its simplest form, cranial kinesis requires a transverse ‘hinge’ across the top or back of the skull, and a sliding basal joint, producing a functional separation between upper jaws and braincase. In modern lizards and snakes, cranial kinesis helps to align the teeth and jaws when grasping prey, and to synchronize

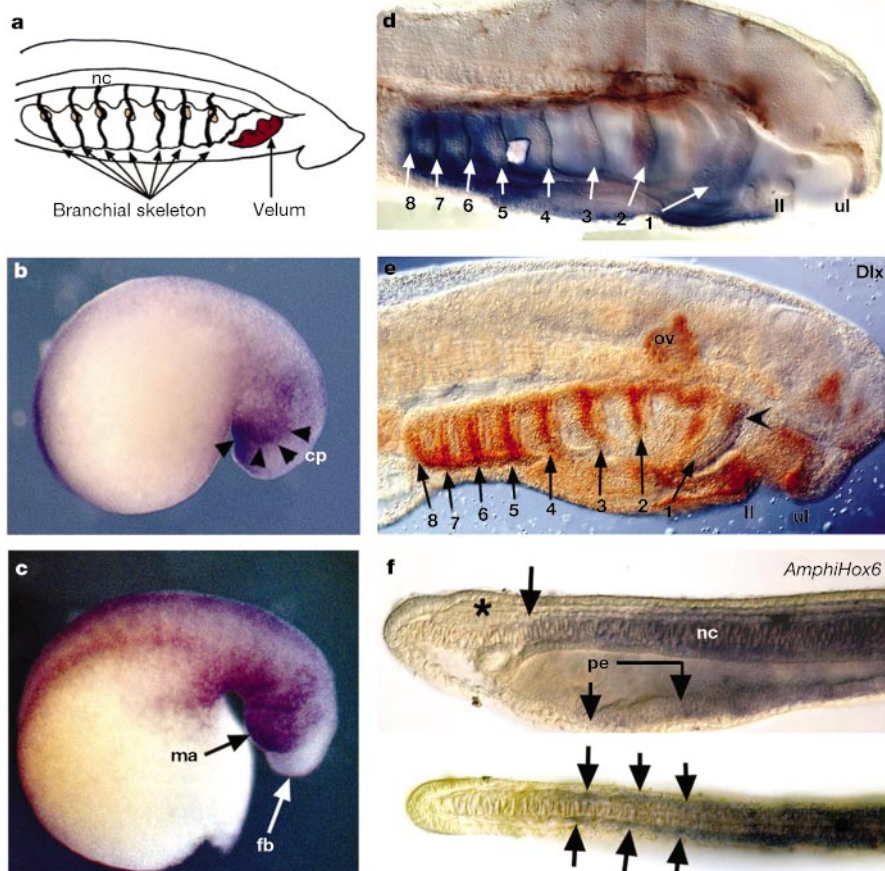


Figure 1 Expression of *Hox6* genes in lamprey and amphioxus embryos. Anterior is to the right in **a–e** and to the left in **f**. **a**, Diagram of the lamprey head. **b–d**, Whole-mount *in situ* hybridization showing *Hox6* expression in lamprey embryos at stage 21 (**b**), 22 (**c**) and 26.5 (**d**). Note the expression in the cheek process (**b**) and mandibular arch (**c**) and the absence of expression from forebrain (**c**). In **d**, *Hox6* expression is shown in pharyngeal arches (numbered). Expression in the mandibular arch (1) is strongest in the ventral region. **e**, Distal-less (*Dlx*) immunostaining of a larval lamprey head. Note the dorsal restriction in the mandibular arch (arrowhead). **f**, Lateral (top) and dorsal (bottom) views of amphioxus early larvae, showing *AmphiHox6* expression. Top arrow, anterior expression boundary in neural tube and notochord, posterior to cerebral vesicle (asterisk); arrows marked ‘pe’, expression boundary in pharyngeal endoderm of first gill slit; paired arrows in bottom image, segmental expression in anterior neural tube. cp, cheek process; ma, mandibular arch; fb, forebrain; ll, lower lip; ul, upper lip; ov, otic vesicle; nc, notochord.

gnathostome chordates and that posterior retraction of *Hox6* expression may have occurred in gnathostomes after their divergence from agnathans.

Comparison of gene expression in lamprey and gnathostome arches has highlighted a surprising conservation between them^{10,11,13}, but has revealed few differences that can account for the evolution of jaws. Given the inhibitory effects of *Hox* genes on jaw formation, loss of *Hox* expression from the first arch and the associated neural crest of early gnathostomes may have facilitated ventral chondrification of the first arch crest, and thus formation of ventral mandibular cartilage.

Previous work proposed that jaws originated either by modification of pre-existing gill arches or by augmentation of a dorsal mandibular structure that resembled the velum of osteostrachans and modern lampreys⁴. My results identify a potential developmental mechanism for the latter hypothesis, and raise the possibility that the ventral mandibular

skeleton was added onto an evolutionarily ancient, velar-like cartilage after *Hox* expression was eliminated from the first pharyngeal arch.

Martin J. Cohn

Division of Zoology, School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK
e-mail: m.j.cohn@reading.ac.uk

- Alexandre, D. *et al.* *Development* **122**, 735–746 (1996).
- Pasqualetti, M. *et al.* *Development* **127**, 5367–5378 (2000).
- Grammatopoulos, G. A. *et al.* *Development* **127**, 5355–5365 (2000).
- Janvier, P. *Early Vertebrates* (Oxford Scientific, New York, 1996).
- Trainor, P. A. & Krumlauf, R. *Nature Rev. Neurosci.* **1**, 116–124 (2000).
- Schilling, T. F. *et al.* *Dev. Biol.* **231**, 201–216 (2001).
- Langille, R. M. & Hall, B. K. *Development* **102**, 301–310 (1988).
- Newth, D. R. *J. Exp. Biol.* **28**, 247–260 (1951).
- Horigome, N. *et al.* *Dev. Biol.* **207**, 287–308 (1999).
- Neideet, A. H. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 1665–1670 (2001).
- Myojin, M. *et al.* *J. Exp. Zool.* **291**, 68–84 (2001).
- Wada, H. *et al.* *Dev. Biol.* **213**, 131–141 (1999).
- Ogasawara, M. *et al.* *Dev. Biol.* **223**, 399–410 (2000).

Supplementary information accompanies this communication on Nature’s website (www.nature.com).

Competing financial interests: declared none.

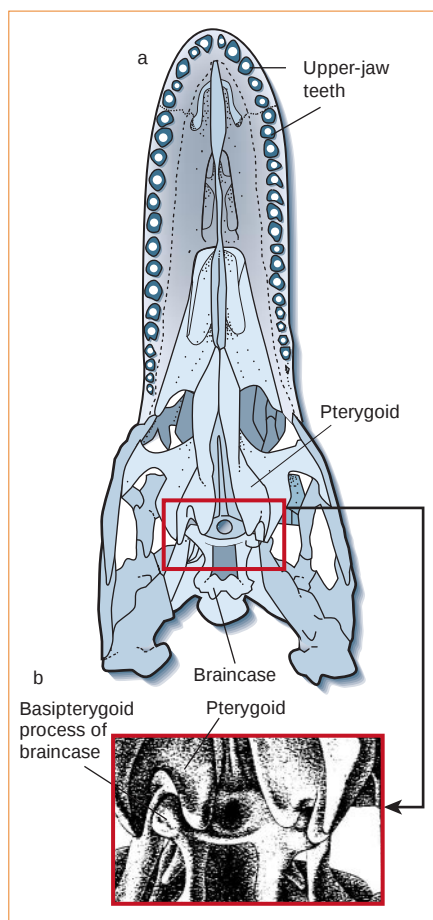


Figure 1 Skull of *Allosaurus fragilis* (after ref. 7). **a**, Ventral view. **b**, Detail of the basal joint between the twin basipterygoid processes of the braincase and the paired pterygoid bones of the palate. Note that the palate is part of a complex that involves the upper jaw bones as well as the side and roofing bones of the skull, the strength of which is unexpectedly high in the model proposed by Rayfield *et al.*⁴.

closure of the jaws onto active quarry^{6,9}. These modern animals have a complex form of kinesis that is related to their loss of the lower temporal bar of the skull. This loss is not necessary for the less complex kinetic movements of many other tetrapods¹⁰.

A hatchet-like attack would bring the anterior axial skeleton — including the braincase — forcefully down upon the prey, with the basal joint intervening between the braincase and jaws. This flexible synovial joint would diminish the penetrating force of the lunge transmitted through the upper teeth, while creating a large stress load on the basal joint itself. However, the basal joint in *Allosaurus* is not unusually robust or reinforced.

If *Allosaurus* used a more usual method of attack, however, the FEA results might be explained by the high torque and shear stress generated by struggling prey held by a very narrow-headed predator (Fig. 1). The allosaur skull has an open design, owing to its large fenestrae and narrow, bony struts. This greater opening of the skull, which weakens it further, requires structural com-

pensation, particularly if the struts are to withstand the pull of the muscles that they support. Hence, the 'overengineering' relative to bite forces may represent a compensation for just such functional factors. A chop/slash interpretation⁴ is therefore questionable on biomechanical grounds, and the FEA results could be explained more simply.

T. H. Frazzetta*, Kenneth V. Kardong†

*Department of Animal Biology/Ecology, Ethology and Evolution, University of Illinois, Urbana, Illinois 61801, USA

†School of Biological Sciences, Washington State University, Pullman, Washington 99164, USA
e-mail: kkardong@wsu.edu

1. Farlow, J. O. *Am. Mid. Nat.* **95**, 186–191 (1976).
2. Farlow, J. O., Brinkman, D. L., Abler, W. L. & Currie, P. J. *Mod. Geol.* **16**, 161–198 (1991).
3. Erickson, G. M. *et al. Nature* **382**, 706–707 (1996).
4. Rayfield, E. J. *et al. Nature* **409**, 1033–1037 (2001).
5. Frazzetta, T. H. *J. Morphol.* **111**, 287–320 (1962).
6. Schwenk, K. in *Feeding in Tetrapod Vertebrates: Form, Function, Phylogeny* (ed. Schwenk, K.) 175–291 (Academic, San Diego, 2000).
7. Madsen, J. H. *J. Utah Geol. Min. Surv. Bull.* **109**, 1–163 (1976).
8. Ahlberg, P. E., Clack, J. A. & Lukševičs, E. *Nature* **381**, 61–64 (1996).
9. Frazzetta, T. H. in *Advances in Herpetology and Evolutionary Biology* (eds Rodin, A. G. J. & Miyata, K.) 222–244 (Harvard Univ. Press, Cambridge, Massachusetts, 1983).
10. Frazzetta, T. H. in *Evolutionary Biology* (eds Hecht, M. K., Wallace, B. & Prance, G. T.) 419–461 (Plenum, New York, 1986).

Rayfield *et al.* reply — Our analysis showed that the cranial strength of *Allosaurus* far exceeded the stresses that could have been generated by its jaw-closing muscles, and we suggested that this disparity might be explained if this dinosaur had adopted a 'slash-and-tear' mode of attack. Frazzetta and Kardong raise three points in criticism of this suggested mode of feeding.

First, they point out that no living tetrapod (lizard) attacks prey in quite this way. However, there are no living equivalents of the 1–2-tonne, bipedal, very large-headed predators. Given that *Allosaurus* was the top predator in the Late Jurassic period¹ and its prey included a range of large, herbivorous dinosaurs (ornithomorphs, stegosaurs and sauropods), its feeding strategies may not have conformed to those of small, modern lizards or even the largest of snakes.

Concerning the tooth and jaw morphology of *Allosaurus*, the upper teeth show significant size variation as they emerge in a staggered tooth-eruption sequence, presenting a discontinuous, coarse profile. This arrangement is ideally suited to creating a deep, lacerating wound. Furthermore, the upper tooth row is longer than the lower and contains more teeth, whereas the lower jaw is relatively slender and tapers towards its tip. These features of the lower jaw are concordant with our interpretation that the upper jaw is the primary weapon of attack.

Assuming that subsequent dismemberment of prey requires lower-jaw dentition, on what basis should the teeth be multi-

regional in *Allosaurus*? Strong 'regionalization' of teeth is a trademark of heterodont synapsids. Also, contrary to the claims of Frazzetta and Kardong, there are significant differences between many theropods in terms of tooth shape, head shape and body size^{2,3}. *Allosaurus* is the only taxon to which FEA has been applied, so it is not currently possible to compare feeding strategies among theropods.

Frazzetta and Kardong argue that cranial flexibility (kinesis) requires a transverse hinge across the top or back of the skull. The skull roof and back of the skull of *Allosaurus fragilis* are not hinged transversely and there is a complete lower temporal bar. The authors note that the basal articulation is flexible and imply that it was involved in fore–aft intracranial sliding, as seen in kinetic lizard skulls. In fact, the basal articulation is not capable of sliding in this sense, but allows the palate to rotate safely against the braincase when the skull is subjected to vertical bending forces⁴.

Furthermore, intracranial displacement between the palatal bones and the cheek region reduces stress at the basal articulation during an impact bite⁴, so 'basal robusticity' at this joint is not required. The conditions required for pro-, meso- or metakinesis are not met in the allosaur skull, and modern lizards and snakes cannot be considered as an appropriate analogue.

Frazzetta and Kardong's comments on our analysis erroneously apply lepidosaur cranial mechanisms to theropod dinosaur skulls. A vertically deep, narrow-headed predator would find its head structurally compromised by high torque when wrestling with large prey⁵. As we pointed out, shear stresses are insignificant and fenestrae strengthen, rather than weaken, the skull when it is subjected to large vertical forces. Recent analysis⁴ shows the lower jaw to be substantially weaker than the upper jaw. 'Slash-and-tear' feeding rather than 'predator–prey struggling' best explains this disparity.

Although our attack strategy was proposed cautiously as an interpretation in the light of the original FEA analysis, we find no evidence here to make us radically revise our initial suggestion.

E. J. Rayfield, D. B. Norman, P. Upchurch

Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK
e-mail: eray@esc.cam.ac.uk

1. Dodson, P., Behrensmeyer, A. K., Baker, R. T. & McIntosh, J. S. *Paleobiology* **6**, 208–232 (1980).
2. Farlow, J. O., Brinkman, D. L., Abler, W. L. & Currie, P. J. *Mod. Geol.* **16**, 161–198 (1991).
3. Abler, W. L. *Paleobiology* **18**, 161–183 (1992).
4. Rayfield, E. J. *Cranial Form and Function in a Large Theropod Dinosaur: A Study Using Finite Element Analysis*. Thesis, Univ. Cambridge (2001).
5. Busbey, A. B. in *Functional Morphology in Vertebrate Paleontology* (ed. Thomason, J. J.) 173–192 (Cambridge Univ. Press, 1995).

Ecological responses to recent climate change

Gian-Reto Walther*, Eric Post†, Peter Convey‡, Annette Menzel§, Camille Parmesan||, Trevor J. C. Beebee¶, Jean-Marc Fromentin#, Ove Hoegh-Guldberg* & Franz Bairlein**

* Institute of Geobotany, University of Hannover, Nienburger Str. 17, 30167 Hannover, Germany

† Department of Biology, The Pennsylvania State University, 208 Mueller Lab, University Park, Pennsylvania 16802, USA

‡ British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK

§ Department of Ecology, Technical University Munich, Am Hochanger 13, 85354 Freising, Germany

|| Integrative Biology, Patterson Labs 141, University of Texas, Austin, Texas 78712, USA

¶ School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

IFREMER, Centre Halieutique Méditerranéen et Tropical, Blvd Jean Monnet, BP 171, 34203 Sète Cedex, France

* Centre for Marine Studies, University of Queensland, St Lucia, 4072 Queensland, Australia

** Institute for Avian Research 'Vogelwarte Helgoland', An der Vogelwarte 21, 26386 Wilhelmshaven, Germany

There is now ample evidence of the ecological impacts of recent climate change, from polar terrestrial to tropical marine environments. The responses of both flora and fauna span an array of ecosystems and organizational hierarchies, from the species to the community levels. Despite continued uncertainty as to community and ecosystem trajectories under global change, our review exposes a coherent pattern of ecological change across systems. Although we are only at an early stage in the projected trends of global warming, ecological responses to recent climate change are already clearly visible.

The Earth's climate has warmed by approximately 0.6 °C over the past 100 years with two main periods of warming, between 1910 and 1945 and from 1976 onwards. The rate of warming during the latter period has been approximately double that of the first and, thus, greater than at any other time during the last 1,000 years¹. Organisms, populations and ecological communities do not, however, respond to approximated global averages. Rather, regional changes, which are highly spatially heterogeneous (Fig. 1), are more relevant in the context of ecological response to climatic change. In many regions there is an asymmetry in the warming that undoubtedly will contribute to heterogeneity in ecological dynamics across systems. Diurnal temperature ranges have decreased because minimum temperatures are increasing at about twice the rate of maximum temperatures. As a consequence, the freeze-free periods in most mid- and high-latitude regions are lengthening and satellite data reveal a 10% decrease in snow cover and ice extent since the late 1960s. Changes in the precipitation regime have also been neither spatially nor temporally uniform (Fig. 1). In the mid- and high latitudes of the Northern Hemisphere a decadal increase of 0.5–1% mostly occurs in autumn and winter whereas, in the sub-tropics, precipitation generally decreases by about 0.3% per decade¹.

There is now ample evidence that these recent climatic changes have affected a broad range of organisms with diverse geographical distributions^{2–6}. We assess these observations using a process-oriented approach and present an integrated synopsis across the major taxonomic groups, covering most of the biomes on Earth. We focus on the consequences of thirty years of warming at the end of the twentieth century, and review the responses in (1) the phenology and physiology of organisms, (2) the range and distribution of species, (3) the composition of and interactions within communities, and (4) the structure and dynamics of ecosystems, highlighting common and contrasting features amongst the taxa and systems considered.

Phenology

Phenology—the timing of seasonal activities of animals and plants—is perhaps the simplest process in which to track changes in the ecology of species in response to climate change. Birds,

butterflies and wild plants, in particular, include popular and easily identifiable species and thus have received considerable attention from the public. As a result many long-term phenological data sets have been collected. Studies in Europe and North America have revealed phenological trends that very probably reflect responses to recent climate change^{7,8}. Common changes in the timing of spring activities include earlier breeding or first singing of birds, earlier arrival of migrant birds, earlier appearance of butterflies, earlier choruses and spawning in amphibians and earlier shooting and flowering of plants (Fig. 2). In general, spring activities have occurred progressively earlier since the 1960s (Table 1).

Some evidence also indicates a later onset of autumnal phenological events, but these shifts are less pronounced and show a more heterogeneous pattern. Studies reveal different proportions of bird species which advance, delay or do not change autumn migration⁹, and trends of leaf colouring of trees at neighbouring stations often show contradictory signals¹⁰. In Europe, for example, leaf colour changes show a progressive delay of 0.3–1.6 days per decade, whereas the length of the growing season has increased in some areas by up to 3.6 days per decade over the past 50 years^{8,11}. This extension of the growing season accords with the lengthening of 12 ± 4 days derived from satellite data¹² as well as with an advance in the seasonal cycle by 7 days and an increase in amplitude of the annual CO₂ cycle since the 1960s¹³.

Environmental links

In contrast with the climatic factors controlling autumn phenology, the climate signal controlling spring phenology is fairly well understood: nearly all phenophases correlate with spring temperatures in the preceding months. For birds, temperatures on the migration route are also important. Some spring events, such as egg-laying of several song birds and the start of the vegetation period in northern and central Europe, also correlate with the North Atlantic Oscillation (NAO) index, which quantifies winter climatic conditions^{5,14,15} (Fig. 2). An analysis of 50 years of data on 13 plant species in 137 localities revealed responses to the NAO in 71% of the total, with early-blooming and herbaceous species showing greater responses to winter warming than late-blooming and woody plants¹⁵. The temperature response of bird arrival may be modified by photo-

periodic control, genetic regulatory systems and/or population size. Phenological changes in birds and plants are often similar, as described in some cross-system studies^{16,17} (Fig. 2). However, the timing of change in different taxonomic groups is not always synchronous and may have profound ecological consequences. Earlier leaf unfolding, for example, generates a longer growing season but may also increase the risk of damage by a late frost (see also 'Complex Dynamics' below).

Variability and inconsistencies

Geographical differences are evident for both plants and birds, with delayed rather than earlier onset of spring phases in southeastern Europe, including later bird arrival in the Slovak Republic¹⁸ and a later start of the growing season in the Balkan region¹¹. Longer data series covering the last century also include periods of later onset¹⁶. There can also be differences in response to climate change between species at particular sites or with time of season. For plants, strong seasonal variation is reported with the highest advances in early spring (and notable advances of succeeding phenophases) and almost no response in summer and early autumn^{10,17}. Similarly, short-distance migrating birds, which tend to migrate early in the season, often exhibit a trend towards earlier arrival, whereas the later arrivals by long-distance migrants show a more complex

response, with many species not changing their arrival times or even delaying them^{7,19}.

Range shifts to keep up with climate change

It is generally agreed that climatic regimes influence species' distributions, often through species-specific physiological thresholds of temperature and precipitation tolerance^{20,21}. With general warming trends, these 'climate envelopes' become shifted towards the poles or higher altitudes. To the extent that dispersal and resource availability allow, species are expected to track the shifting climate and likewise shift their distributions poleward in latitude and upward in elevation. In some cases (for example, reef-building corals), range shifts in response to changing temperature may not occur if latitudinal distributions are also limited by other factors such as light⁴¹.

Many studies of the biological impacts of climate change have focused on species abundances and distributions in search of the predicted systematic shifts. Migratory species are among the best documented but often exhibit large fluctuations from year to year in their breeding sites, making it difficult to discern long-term range shifts^{22,23}. By contrast, range changes in more sedentary species follow from the slow processes of population extinctions and colonizations. This has made it easier to detect true geographic

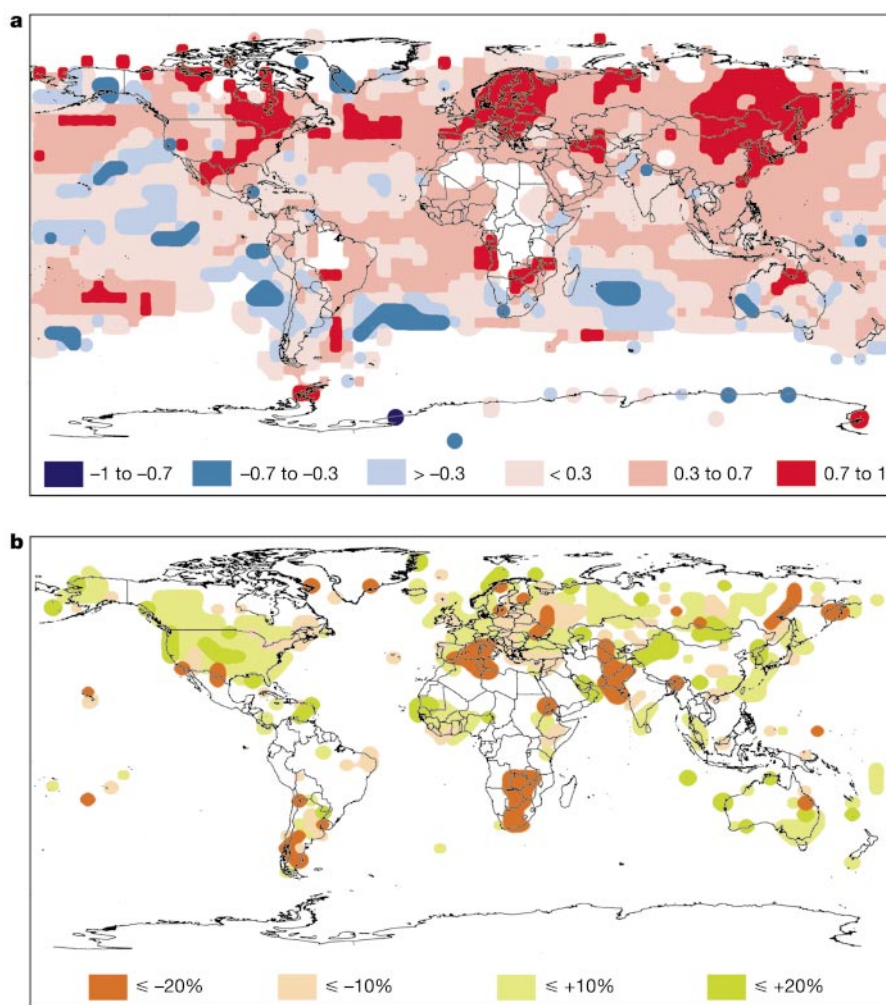


Figure 1 Spatial variability of annual trends in temperature and precipitation since 1976 relative to 1961 to 1990 normals (ref. 1, modified). **a**, Temperature (°C per decade); **b**, precipitation (% per decade).

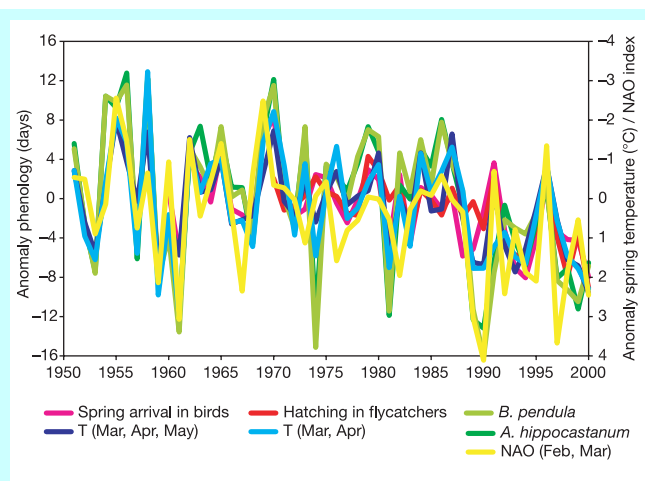


Figure 2 Anomalies of different phenological phases in Germany correlate well with anomalies of mean spring air temperature T and NAO index (by P. D. Jones, <http://www.cru.uea.ac.uk/cru/data/nao.htm>). Temperature taken from 35 German climate stations. Phenological phases used: spring arrival in birds, island of Helgoland, North Sea; hatching in flycatchers (*Ficedula hypoleuca*), Northern Germany; and mean onset of leaf unfolding of *Aesculus hippocastanum* and *Betula pendula*.

shifts in the latter group because change is more methodical and missing data have a smaller impact. It is now clear that poleward and upward shifts of species ranges have occurred across a wide range of taxonomic groups and geographical locations during the twentieth century^{2,4,6,23} (Table 2).

Factors affecting species distribution interact in complex ways, and it is not surprising that simple correlations with temperature changes are not always observed. Range shifts are often episodic rather than gradual or monotonic. In regions under the influence of El Niño/Southern Oscillation (ENSO), for example, change may happen rapidly during warm episodes, with 'setbacks' during cool periods. In addition, climatic extremes—related to natural oscillations and underlying long-term trends—are also important in driving the present range changes²³. Thus, rates of range shifts vary greatly among and within species, implying differential dispersal abilities. Whereas the magnitude of elevational shifts of alpine plant species lags behind the isothermal shift of 8–10 m per decade²⁴, butterflies appear to track decadal warming quickly^{25,26}, matching the upwards and northwards shifts of temperature isotherms²⁷ (compare data in Table 2).

Climate change and invasions

With climate change, non-native species from adjacent areas may cross frontiers and become new elements of the biota. When long distances have been covered, such movements have often been mediated by human activity. However, for species originating from habitats more suitable than the new location provides, a permanent establishment at the new locality may not be possible without changes in local conditions. An obvious possibility is that, while human activities promote species movement, their subsequent reproduction and spread at the new location imply altered site conditions due, for example, to climate change. The clearest

evidence for such a climate trigger occurs where a suite of species with different histories of introduction spread *en masse* during periods of climatic amelioration^{28,29}. Examples include warm-water species that have recently appeared in the Mediterranean and the North seas^{28,30,31} and thermophilous plants that spread from gardens into surrounding countryside^{29,32} (Fig. 3). Even in such remote places as some sub-Antarctic islands, it is estimated that introductions by humans over the last two centuries account for 50% or more of the higher-plant diversity³³ and a considerable proportion of the insect and mite faunas³⁴.

Climate-linked invasions might also involve the immigration of unwanted neighbours such as epidemic diseases. There is much evidence that a steady rise in annual temperatures has been associated with expanding mosquito-borne diseases in the highlands of Asia, East Africa and Latin America³⁵. Overall, trends of range changes show remarkable internal consistency between studies relating to glaciers, plant and insect ranges and shifting isotherms.

Community shifts

The assemblages of species in ecological communities reflect interactions among organisms as well as between organisms and the abiotic environment. We might expect, therefore, that rapid climatic change or extreme climatic events can alter community composition. In the Sonoran desert of the southwestern United States, for example, recent increases in woody shrub density, extinction of previously common animal species and increases in formerly rare animal species have been attributed to regional climatic shifts³⁶. Furthermore, some of the examples of range shifts mentioned earlier involve community-level changes^{29,37,38}. Changes in distribution are often asymmetrical with species invading faster from lower elevations³⁸ or latitudes³⁷ than resident species are receding upslope or poleward. The result is a (presumably transient) increase in species richness of the community in question as a consequence of the variability in rates at which species shift their ranges.

Observed changes at extremes of environmental and biological gradients

Rapid environmental warming has been reported over the last 30–50 years at a number of stations in the Antarctic, particularly in the Antarctic Peninsula region and on sub-Antarctic islands, along with changes in precipitation patterns^{39,40}. Likewise, tropical oceans have increased in temperature by 1–2°C over the past 100 years⁴¹. Warming trends are punctuated in most oceans by fairly regular phenomena such as ENSO events, which induce anomalies both locally and temporarily; these phenomena have increased in size and duration over the past century⁴². Most climate projections reveal that this trend is likely to increase rapidly in the next 50 years⁴¹.

Antarctic terrestrial habitats and nearshore tropical marine communities reflect opposite extremes of environmental and biological variation on Earth. The warmth and stability of tropical oceans contrasts dramatically with the year-round low temperatures and rapid, unpredictable short-term variation typical on land in the Antarctic. Likewise, the rich biodiversity and trophic web complexity characterizing reef communities differ strikingly from the characteristics of the extremely simple communities of the Antarctic (Fig. 4). Both environments are experiencing extensive changes^{43,44}

Table 1 Evidence of recent advances in the timing of spring events

Taxon	Location	Observed changes	Period
Numerous plant species	Europe North America	Flowering and leaf unfolding occurring 1.4–3.1 days per decade earlier ⁸ Flowering and leaf unfolding occurring 1.2–2.0 (3.8) days per decade earlier ⁸	Past 30–48 years Past 35–63 years
18 butterfly species	UK	Earlier appearance by 2.8–3.2 days per decade ⁶²	Past 23 years
Amphibians	UK	Earlier breeding ⁶⁷	Past 25 years
Numerous bird species	Europe, North America	Earlier spring migration by 1.3–4.4 days per decade and breeding by 1.9–4.8 days per decade ^{7,22,83–86}	Past 30–60 years

Table 2 Recent latitudinal and altitudinal range shifts

Species*	Location	Observed changes	Climate link
Treeline	Europe, New Zealand	Advancement towards higher altitudes ^{87–89}	General warming
Arctic shrub vegetation	Alaska	Expansion of shrubs in previously shrub-free areas ⁹⁰	Environmental warming
Alpine plants	European Alps	Elevational shift of 1–4 m per decade ²⁴	General warming
Antarctic plants and invertebrates	Antarctica	Distribution changes ⁹¹	Liquid water availability and increased temperature
Zooplankton, intertidal invertebrate and fish communities	Californian coast, North Atlantic	Increasing abundance of warm-water species ^{92,97,93,94}	Warmer shoreline ocean temperature
39 butterfly species	North America and Europe	Northward range shifts up to 200 km over 27 years ^{25,95}	Increased temperatures
such as Edith's Checkerspot butterfly (<i>Euphydryas editha</i>)	Western United States	124 m upward and 92 km northward shift since the beginning of the twentieth century ^{25,26}	
Lowland birds	Costa Rica	Extension of distribution from lower mountain slopes to higher areas ³⁸	Dry season mist frequency
12 bird species	Britain	18.9 km average range movement northwards over a 20-year period ⁹⁶	Winter temperatures
Red fox (<i>Vulpes vulpes</i>), Arctic fox (<i>Alopex lagopus</i>)	Canada	Northward expansion of red fox range and simultaneous retreat of Arctic fox range ⁹⁷	General warming

* Where possible, numbers of species which showed a response to climate change are given.

and, lying at opposite extremes of environmental and biological gradients, generate different biological signals of climate change.

In Antarctic terrestrial ecosystems, visually dramatic examples of biological changes in response to climatic warming include the colonization by macroscopic plants (largely mosses) of previously bare or newly exposed ground and the rapid expansion in extent and numbers of the only two higher plants present on the continent^{43,45}. Commensurate with vegetational changes has been colonization by soil invertebrates. As yet there are few examples of Antarctic colonization by 'exotics' from lower latitudes, although some species have established themselves rapidly around several sources of geothermal warming⁴⁶ and there is an increasing number of human-mediated imports, particularly to the sub-Antarctic region.

Substantial impacts on community structure have been observed in coral reefs during periods of warmer than normal sea temperatures. Poised near their upper thermal limits, coral reefs have undergone global mass bleaching events whenever sea temperatures have exceeded long-term summer averages by more than 1.0 °C for several weeks^{41,47}. Six periods of mass coral bleaching have occurred

since 1979 and the incidence of mass coral bleaching is increasing in both frequency and intensity⁴¹. The most severe period occurred in 1998, in which an estimated 16% of the world's reef-building corals died⁴⁸. The impact of thermal stress on reefs can be dramatic, with the almost total removal of corals in some instances^{41,49–51}. In some cases (usually smaller or shorter thermal anomalies), thin-tissued, branching acroporid and pocilloporid corals have bleached and/or died preferentially, leaving more massive species like *Porites* spp intact. In other cases, all coral species have been largely removed^{51,52}. Estimates of how ecosystem species richness and community structure have changed after bleaching events are generally unavailable, but such changes are suspected to be large.

The combination of rising temperatures and ENSO variability has generated contrasting impacts in Antarctic terrestrial versus tropical marine ecosystems. Increasing temperatures have reduced the likelihood that Antarctic organisms will be exposed to their lower thermal limits, thereby allowing increases in both numbers and extent of populations previously at the edge of their range while also, in a few instances, increasing the risk of exposure to upper

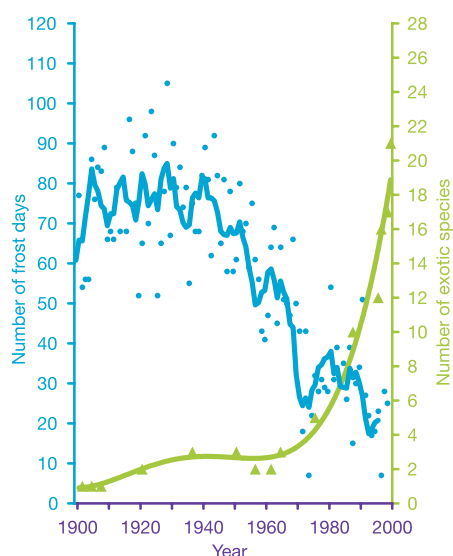


Figure 3 Vegetation shift from indigenous deciduous to exotic evergreen broad-leaved vegetation in southern Switzerland. The shrub layer is dominated by the growing number of spreading exotic evergreen broad-leaved species (see illustration) that

appear to profit from milder winter conditions, indicated here by the decreasing number of days with frost per year (the smoothed curve gives five year averages for the number of frost days per year)²⁹.

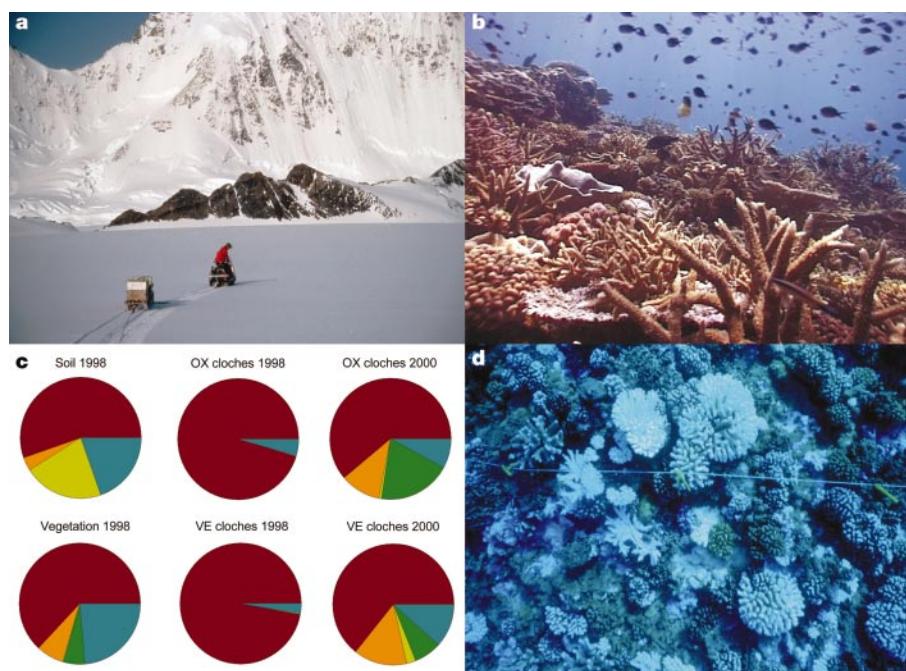


Figure 4 Different environments and their responses to warming. Biodiversity is low in terrestrial communities in Antarctica (**a**), but very high in nearshore marine communities in tropical oceans (**b**). The response of these communities to warming seems to differ. Manipulations of Antarctic soils by using Perspex cloches provided realistic simulations of climate change (temperature increase and exposure to different wavelengths of UV

radiation) by using different Perspex types, VE and OX, between 1998 and 2000. These led within 2 years to an increased diversity in soil nematode communities (colours indicate the proportional contribution of major genera), comparable to that found naturally in more developed microhabitats (**c**). However the increased occurrence of bleaching events on coral reefs (**d**) is likely to decrease abundance if not diversity.

thermal limits^{43,53,54}. In the tropics, anomalies of less than 1 °C may exceed physiological tolerances and result in large-scale coral bleaching (owing to physiological dysfunction and loss of crucial dinoflagellate symbionts) and subsequent mortality. Because of their need to cope with short-term and seasonal environmental variability, Antarctic biota generally occupy a wider physical niche than do organisms from more stable environments (such as marine tropical organisms). This difference is important in defining range boundaries and how species respond to environmental change⁵⁵.

Complex dynamics

Responses by individual species to climate change may disrupt their interactions with others at the same or adjacent trophic levels. When closely interacting or competing species display divergent responses or susceptibilities to change, the outcome of their interactions may be altered, as long-term data on both terrestrial and marine organisms indicate^{56,57}.

Recruitment success and trophic interactions in marine systems

Recruitment in fish populations has long been known to be a key process that is strongly influenced by climate variability⁵⁸. Variations in atmospheric circulation over the Bering Sea, through interactions with ocean currents, influence transportation of juvenile walleye pollock (*Theragra chalcogramma*) away from adults, affecting the intensity of cannibalism and, consequently, year class strength⁵⁹. Because walleye pollock is an important forage species for other fish, marine mammals and birds, its fluctuations in recruitment affect the whole Bering Sea food web. A Southern Ocean parallel involves krill (*Euphausia superba*), a key food source for higher predators (penguins and other seabirds, whales, seals) as well as a fishery target. Climate change is apparently affecting the reproductive grounds of krill, and consequently its recruitment, by reducing the area of sea ice formed near the Antarctic

Peninsula, which leads to both food web and human economic consequences^{58,60}.

The most widespread effects of climate on dynamics in marine systems appear, however, to be indirect. The persistence of positive anomalies of the North Atlantic Oscillation has, for instance, modified marine primary and secondary production⁵⁷. This may affect the availability of planktonic food for fish larvae, which determines the recruitment success and consequently the size of fish populations⁵⁸. Migration patterns and spatial distributions of large pelagic fish, such as bluefin tuna (*Thunnus thynnus*), can also be altered indirectly through climate-induced changes in prey abundance⁶¹. In upwelling systems, fish production appears to be controlled by enrichment, concentration and retention processes⁶², which are themselves governed by climatic factors. Because temperature increases should intensify upwelling, global fish production could decline because of a consequent reduction in the concentration and retention processes. Changes in the north-east Pacific ecosystem that support this hypothesis are already evident⁶³.

Human exploitation may further exacerbate the effects of oceanic warming on fish populations. In North Sea cod (*Gadus morhua*), for example, a long adult lifespan provides a buffer to occasional recruitment failures, but overfishing has truncated the age structure of the population and thereby increased vulnerability to the adverse effects of prolonged warming⁶⁴.

Species interactions in terrestrial systems

Direct climatic effects on development, spatial distribution, and species interactions are apparent in amphibians and reptiles, which, in common with other ectotherms, are heavily influenced by environmental conditions. Both temperature and humidity affect their reproductive physiology and population dynamics. Oogenesis and spermatogenesis in temperate amphibians and reptiles are dependent on seasonal temperature regimes. In the case of reptiles

there is particular interest in the effects of climate change on the population dynamics of species with temperature-dependent sex determination. In painted turtles (*Chrysemys picta*) offspring sex ratio is highly correlated with mean July temperature, and the production of male offspring would be potentially compromised even by modest (2–4 °C) temperature increases^{65,66}.

Winter warming has precipitated breeding season changes in some but not all species of amphibians in Britain⁶⁷. This variability has, in turn, altered temporal niche overlaps in breeding ponds with immediate consequences for trophic interactions. Thus, newts (*Triturus* spp.) are entering ponds earlier than before, whereas frogs (*Rana temporaria*) have not substantially altered their reproductive phenology. Embryos and larvae of early-breeding frogs are, therefore, exposed to higher levels of newt predation. Such examples illustrate the higher-order consequences of phenological responses to climate change described above. Especially dramatic indirect effects have been observed on the population dynamics of montane amphibian species³⁸. In Costa Rica and the western USA, sharp population declines have been linked with epidemic disease and changes in precipitation patterns related to recent warming^{38,68}. Further, frog population declines in Costa Rica have occurred simultaneously with those of anoline lizards (*Norops* spp.), both being associated with the same climatic patterns³⁸.

Delays in spring arrival by migratory birds may lead to increased competition for nest sites with species arriving earlier⁶⁹. Evidence also indicates that warmer spring weather in Europe has disrupted the synchrony between winter moth (*Operophtera brumata*) hatching and oak bud burst, leading to a mismatch between the peak in insect availability and the peak food demands of great tit (*Parus major*) nestlings^{70,71}. Such disharmonization of fine-tuned events may pose consequences for species interactions and the persistence of ecological communities across an array of ecosystems.

Extensive studies of large mammals indicate that climatic extremes appear to influence juvenile survival, primarily during winter, although not independently of population density^{15,72}. Increasingly warm winters associated with the NAO influence the development and fecundity of red deer (*Cervus elaphus*)⁷³ and Soay sheep (*Ovis aries*)⁷⁴ in Norway and the UK. The impact of such life history responses on population dynamics can occur years later when cohorts have reached reproductive maturity^{15,74} and may, as in the case of Soay sheep, occur only above certain population densities⁷⁵. On Isle Royale, USA, climate directly influences temporal dynamics at producer, herbivore and carnivore trophic levels⁷⁶, as well as indirectly through mediation of trophic interactions such as wolf predation and moose herbivory⁷⁷.

Knowns and unknowns

We have reviewed merely a portion of the enormous body of basic research on ecological and physiological processes that are sensitive to climatic variables such as temperature and precipitation. The evidence indicates that only 30 years of warmer temperatures at the end of the twentieth century have affected the phenology of organisms, the range and distribution of species, and the composition and dynamics of communities. These examples, spanning the previous century and encompassing most major taxa and ecosystems on Earth, provide linkages between recently observed changes in natural systems and twentieth century climate change. The mechanistic bases for the observed biotic responses to climate change have been well established through experimental and observational studies on the behaviour, ecology and physiology of many wild species. Such studies will continue to provide detailed mechanisms by which climatic change affects individual physiology, seasonal timing, population dynamics and geographic distributions.

However, the complexity of ecological interactions renders it difficult to extrapolate from studies of individuals and populations to the community or ecosystem level. We do not, for example, have a

clear understanding of the roles of short-term versus long-term environmental stochasticity and population-intrinsic processes in community dynamics and stability^{76,78}. Currently, the most relevant physical and temporal scales of ecological investigation are local and short-term (less than three decades). In contrast, climatology generally encompasses much larger spatial and temporal scales. As a consequence, it remains difficult to link population and community-level dynamics to the global-scale studies of atmospheric and oceanic processes^{23,79,80}.

As both ecological theory and conservation history have shown, the modern landscape provides little flexibility for ecosystems to adjust to rapid environmental changes. In contrast with historical responses and migration processes, species in many areas today must move through a landscape that human activity has rendered increasingly impassable⁸¹. As a result of the widespread loss and fragmentation of habitats, many areas which may become climatically suitable with future warming are remote from current distributions, and beyond the dispersal capacity of many species. Consequently, species with low adaptability and/or dispersal capacity will be caught by the dilemma of climate-forced range change and low likelihood of finding distant habitats to colonize, ultimately resulting in increased extinction rates. Furthermore, several case studies (especially in the marine environment) have indicated that climate change can reinforce the detrimental effects of human exploitation and mismanagement and push species and ecosystem tolerances over their limits. This is exemplified by the North Sea cod, which clearly illustrates the human economic consequences of such synergistic effects, as well as by the massive and direct impacts of climate change on coral reefs that may yield even more substantial social and economic impacts.

It is not simply the magnitude of change of global average temperature over the last century but the inherent asymmetry in change processes that complicates predictions of ecological responses, especially for complex systems. However, it is clear that communities are already undergoing re-assembly that is attributable to climate change, as several of the studies cited in this review demonstrate. The implications of such large-scale, consistent responses to relatively low average rates of climate change are large and the projected warming for the coming decades raises even more concern about its ecological and also socio-economic consequences. □

1. Climate Change 2001. Third Assessment Report of the Intergovernmental Panel on Climate Change IPCC (WG I & II) (Cambridge Univ. Press, Cambridge, 2001).
2. Hughes, L. Biological consequences of global warming: is the signal already apparent? *Trends Ecol. Evol.* **15**, 56–61 (2000).
3. Wuethrich, B. How climate change alters rhythms of the wild. *Science* **287**, 793–795 (2000).
4. McCarthy, J. P. Ecological consequences of recent climate change. *Conserv. Biol.* **15**(2), 320–331 (2001).
5. Ottersen, G. et al. Ecological effects of the North Atlantic Oscillation. *Oecologia* **128**, 1–14 (2001).
6. Walther, G.-R., Burga, C. A. & Edwards, P. J. (eds) "Fingerprints" of Climate Change—Adapted Behaviour and Shifting Species Ranges (Kluwer Academic/Plenum, New York, 2001).
7. Bairlein, F. & Winkel, W. in *Climate of the 21st Century: Changes and Risks* (eds Lozan, J. L., Graßl, H. & Hupfer, P.) 278–282 (Wissenschaftliche Auswertungen, Hamburg, 2001).
8. Menzel, A. & Estrella, N. in "Fingerprints" of Climate Change—Adapted Behaviour and Shifting Species Ranges (eds Walther, G.-R., Burga, C. A. & Edwards, P. J.) 123–137 (Kluwer Academic, New York, 2001).
9. Gatter, W. Timing and patterns of visible autumn migration: Can effects of global warming be detected? *J. Ornithol.* **133**(4), 427–436 (1992).
10. Menzel, A., Estrella, N. & Fabian, P. Spatial and temporal variability of the phenological seasons in Germany from 1951–1996. *Glob. Change Biol.* **7**, 657–666 (2001).
11. Menzel, A. & Fabian, P. Growing season extended in Europe. *Nature* **397**, 659 (1999).
12. Myrneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).
13. Keeling, C. D., Chin, J. F. S. & Whorf, T. P. Increased activity of northern vegetation inferred from atmospheric CO₂ measurements. *Nature* **382**, 146–149 (1996).
14. Forchhammer, M. C., Post, E. & Stenseth, N. C. Breeding phenology and climate. *Nature* **391**, 29–30 (1998).
15. Post, E. & Stenseth, N. C. Climatic variability, plant phenology, and northern ungulates. *Ecology* **80**, 1322–1339 (1999).
16. Ahas, R. Long-term phyto-, ornitho- and ichthyophenological time-series analyses in Estonia. *Int. J. Biometeorol.* **42**, 119–123 (1999).
17. Bradley, N. L., Leopold, A. C., Ross, J. & Huffaker, W. Phenological changes reflect climate change in Wisconsin. *Proc. Natl Acad. Sci. USA* **96**, 9701–9704 (1999).
18. Sparks, T., Heyen, H., Braslavskaya, O. & Lehtikoinen, E. Are European birds migrating earlier? *BTO News* **223**, 8–9 (1999).

19. Huin, N. & Sparks, T. H. Spring arrival patterns of the cuckoo *Cuculus canorus*, nightingale *Luscinia megarhynchos* and spotted flycatcher *Muscicapa striata* in Britain. *Bird Study* **47**, 22–31 (2000).
20. Hoffman, A. A. & Parsons, P. A. *Extreme Environmental Change and Evolution* (Cambridge Univ. Press, Cambridge, 1997).
21. Woodward, F. I. *Climate and Plant Distribution* (Cambridge Univ. Press, Cambridge, 1987).
22. Crick, H. Q. P. & Sparks, T. H. Climate change related to egg-laying trends. *Nature* **399**, 423–424 (1999).
23. Easterling, D. R. *et al.* Climate extremes: observations, modeling, and impacts. *Science* **289**, 2068–2074 (2000).
24. Grabherr, G., Gottfried, M. & Pauli, H. Climate effects on mountain plants. *Nature* **369**, 448 (1994).
25. Parmesan, C. *et al.* Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* **399**, 579–583 (1999).
26. Parmesan, C. Climate and species' range. *Nature* **382**, 765–766 (1996).
27. Karl, T. R., Knight, R. W., Easterling, D. R. & Quayle, R. G. Indices of climate change for the United States. *Bull. Am. Meteorol. Soc.* **77**, 279–292 (1996).
28. Nehring, S. Establishment of thermophilic phytoplankton species in the North Sea: biological indicators of climatic changes? *ICES J. Mar. Sci.* **55**, 818–823 (1996).
29. Walther, G.-R. Climatic forcing on the dispersal of exotic species. *Phytocoenologia* **30**(3–4), 409–430 (2000).
30. Chisholm, J. R. M., Jaubert, J. M. & Giaccone, G. *Caulerpa taxifolia* in the northwest Mediterranean: introduced species or migrant for the Red Sea? *C. R. Acad. Sci. Life Sci.* **318**, 1219–1226 (1995).
31. Nieder, J., La Mesa, G. & Vacchi, M. Blenniidae along the Italian coasts of the Ligurian and the Tyrrhenian Sea: Community structure and new records of *Scartella cristata* for northern Italy. *Cybiurn* **24**(4), 359–369 (2000).
32. Dukes, J. S. & Mooney, H. A. Does global change increase the success of biological invaders? *Trends Ecol. Evol.* **14**(4), 135–139 (1999).
33. Smith, R. I. L. Introduced plants in Antarctica: potential impacts and conservation issues. *Biol. Conserv.* **76**, 135–146 (1996).
34. Pugh, P. J. A. Non-indigenous Acari of Antarctica and the sub-Antarctic islands. *Zool. J. Linn. Soc.* **110**, 207–217 (1994).
35. Epstein, P. R. *et al.* Biological and physical signs of climate change: focus on mosquito-borne diseases. *Bull. Am. Meteorol. Soc.* **79**(3), 409–417 (1998).
36. Brown, J. H., Valone, T. J. & Curtin, C. G. Reorganization of an arid ecosystem in response to recent climate change. *Proc. Natl Acad. Sci. USA* **94**, 9729–9733 (1997).
37. Sagarin, R. D., Barry, J. P., Gilman, S. E. & Baxter, C. H. Climate-related change in an intertidal community over short and long time scales. *Ecol. Monogr.* **69**, 465–490 (1999).
38. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. Biological response to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
39. King, J. C. & Harangozo, S. A. Climate change in the western Antarctic peninsula since 1945: observations and possible causes. *Ann. Glaciol.* **27**, 571–575 (1998).
40. Bergstrom, D. M. & Chown, S. L. Life at the front: history, ecology and change on southern ocean islands. *Trends Ecol. Evol.* **14**, 472–476 (1999).
41. Hoegh-Guldberg, O. Climate change, coral bleaching and the future of the world's coral reefs. *Mar. Freshwat. Res.* **50**, 839–866 (1999).
42. Reaser, J. K., Pomeroy, R., Thomas, P. O. Coral bleaching and global climate change: scientific findings and policy recommendations. *Conserv. Biol.* **14**, 1500–1511 (2000).
43. Convey, P. in "Fingerprints" of Climate Change—Adapted Behaviour and Shifting Species Ranges (eds Walther, G.-R., Burga, C. A. & Edwards, P. J.) 17–42 (Kluwer Academic, New York, 2001).
44. Bryant, D., Burke, L., McManus, J. & Spalding, M. *Reefs at Risk: a Map-based Indicator of Threats to the World's Coral Reefs* (World Resources Institute, Washington DC, 1998).
45. Smith, R. I. L. Plant colonisation response to climate change in the Antarctic. *Folia Fac. Sci. Nat. Univ. Masarykianae Brunensis, Geographia* **25**, 19–33 (2001).
46. Convey, P., Smith, R. I. L., Hodgson, D. A. & Peat, H. J. The flora of the South Sandwich Islands, with particular reference to the influence of geothermal heating. *J. Biogeogr.* **27**, 1279–1295 (2000).
47. Glynn, P. W. Coral reef bleaching in the 1980s and possible connections with global warming. *Trends Ecol. Evol.* **6**, 175–179 (1991).
48. Wilkinson, C. R. (ed.) *Status of Coral Reefs of the World: 2000* (Global Coral Reef Monitoring Network, Australian Institute of Marine Science, Townsville, Australia, 2000).
49. Brown, B. E. Coral bleaching: causes and consequences. *Coral Reefs* **16**, 129–138 (1997).
50. Spencer, T., Teleki, T. A., Bradshaw, C. & Spalding, M. Coral bleaching in the southern Seychelles during the 1997–1998 Indian Ocean warm event. *Mar. Poll. Bull.* **40**, 569–586 (2000).
51. Mumby, P. J. *et al.* Unprecedented bleaching-induced mortality in *Porites* spp. at Rangiroa Atoll, French Polynesia. *Mar. Biol.* **139**(1), 183–189 (2001).
52. Loya, Y. *et al.* Coral bleaching: the winners and the losers. *Ecol. Lett.* **4**, 122–131 (2001).
53. Cunningham, D. M. & Moors, P. J. The decline of rockhopper penguins *Eudyptes chrysocome* at Campbell Island, Southern Ocean and the influence of rising sea temperatures. *Emu* **94**, 27–36 (1994).
54. Barbraud, C. & Weimerskirch, H. Emperor penguins and climate change. *Nature* **411**, 183–186 (2001).
55. Chown, S. L. & Gaston, K. J. Exploring links between physiology and ecology at macro-scales: the role of respiratory metabolism in insects. *Biol. Rev.* **74**, 87–120 (1999).
56. Saetre, G.-P., Post, E. & Kral, M. Can environmental fluctuation prevent competitive exclusion in sympatric flycatchers? *Proc. R. Soc. Lond. B* **266**, 1247–1251 (1999).
57. Fromentin, J.-M. & Planque, B. *Calanus* and environment in the eastern North Atlantic. 2. Influence of the North Atlantic Oscillation on *C. finmarchicus* and *C. helgolandicus*. *Mar. Ecol. Prog. Ser.* **134**, 111–118 (1996).
58. Cushing, D. H. *Population Production and Regulation in the Sea: a Fisheries Perspective* (Cambridge Univ. Press, Cambridge, 1995).
59. Weststad, V., Fritz, L. W., Ingraham, J. W. & Megrey, B. A. On the relationship between cannibalism, climate variability, physical transport, and recruitment of Bering sea walleye pollock (*Theragra chalcogramma*). *ICES J. Mar. Sci.* **57**, 272–278 (2000).
60. Loeb, V. *et al.* Effects of sea-ice extent and krill or salp dominance on the Antarctic food web. *Nature* **387**, 897–900 (1997).
61. Polovina, J. J. Decadal variation in the trans-Pacific migration of northern bluefin tuna (*Thunnus thynnus*) coherent with climate-induced change in prey abundance. *Fish. Oceanogr.* **5**, 114–119 (1996).
62. Bakun, A. in *Reinventing Fisheries Management* (eds Pitcher, T. J., Hart, P. J. B. & Pauly, D.) 331–358 (Kluwer Academic, Dordrecht, 1998).
63. McGowan, J. A., Cayan, D. R. & Dorman, L. M. Climate-ocean variability and ecosystem response in the Northeast Pacific. *Science* **281**, 210–217 (1998).
64. Planque, B. & Frédou, T. Temperature and the recruitment of Atlantic cod (*Gadus morhua*). *Can. J. Fish. Aquat. Sci.* **56**, 2069–2077 (1999).
65. Janzen, F. J. Climate-change and temperature-dependent sex determination in reptiles. *Proc. Natl Acad. Sci. USA* **91**, 7484–7490 (1994).
66. Rhen, T. & Lang, J. W. Among-family variation for environmental sex determination in reptiles. *Evolution* **52**, 1514–1520 (1998).
67. Beebe, T. J. C. Amphibian breeding and climate change. *Nature* **374**, 219–220 (1995).
68. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. Complex causes of amphibian population declines. *Nature* **410**, 681–684 (2001).
69. Both, C. & Visser, M. E. Adjustment to climate change is constrained by arrival date in a long-distance migrant bird. *Nature* **411**, 296–298 (2001).
70. Visser, M. E. & Holleman, L. J. M. Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proc. R. Soc. Lond. B* **268**, 289–294 (2001).
71. Visser, M. E., van Noordwijk, A. J., Tinbergen, J. M. & Lessels, C. M. Warmer springs lead to mistimed reproduction in great tits (*Parus major*). *Proc. R. Soc. Lond. B* **265**, 1867–1870 (1998).
72. Milner, J. M., Elston, D. A. & Albon, S. D. Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *J. Anim. Ecol.* **68**, 1235–1247 (1999).
73. Post, E., Stenseth, N. C., Langvatn, R. & Fromentin, J.-M. Global climate change and phenotypic variation among red deer cohorts. *Proc. R. Soc. Lond. B* **264**, 1317–1324 (1997).
74. Forchhammer, M. C., Clutton-Brock, T. H., Lindström, J. & Albon, S. D. Climate and population density induce long-term cohort variation in a northern ungulate. *J. Anim. Ecol.* **70**, 721–729 (2001).
75. Grenfell, B. T. *et al.* Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674–677 (1998).
76. Post, E. & Forchhammer, M. C. Pervasive influence of large-scale climate in the dynamics of a terrestrial vertebrate community. *BMC Ecol.* **1**, 5 (2001).
77. Post, E., Peterson, R. O., Stenseth, N. C. & McLaren, B. E. Ecosystem consequences of wolf behavioural response to climate. *Nature* **401**, 905–907 (1999).
78. Pascual, M. & Levin, S. A. From individuals to population densities: searching for the intermediate scale of nontrivial determinism. *Ecology* **80**, 2225–2236 (1999).
79. Root, T. L. & Schneider, S. H. Can large-scale climatic models be linked with multiscale ecological studies? *Conserv. Biol.* **7**, 256–270 (1993).
80. Harrington, R., Woivod, I. & Sparks, T. Climate change and trophic interactions. *Trends Ecol. Evol.* **14**, 146–150 (1999).
81. Pitelka, L. F. & The Plant Migration Workshop Group. Plant migration and climate change. *Am. Sci.* **85**, 464–473 (1997).
82. Roy, D. B. & Sparks, T. H. Phenology of British butterflies and climate change. *Glob. Change Biol.* **6**, 407–416 (2000).
83. Crick, H. Q. P., Dudley, C., Glue, D. E. & Thomson, D. L. UK birds are laying eggs earlier. *Nature* **388**, 526 (1997).
84. Brown, J. L., Li, S.-H. & Bhagabati, N. Long-term trend toward earlier breeding in an American bird: A response to global warming? *Proc. Natl Acad. Sci. USA* **96**, 5565–5569 (1999).
85. Dunn, P. O. & Winkler, D. W. Climate change has affected the breeding date of tree swallows throughout North America. *Proc. R. Soc. Lond. B* **266**, 2487–2490 (1999).
86. Inouye, D. W., Barr, B., Armitage, K. B. & Inouye, B. D. Climate change is affecting altitudinal migrants and hibernating species. *Proc. Natl Acad. Sci. USA* **97**, 1630–1633 (2000).
87. Kullman, L. 20th century climate warming and tree-limit rise in the southern Scandes of Sweden. *Ambio* **30**(2), 72–80 (2001).
88. Meshinev, T., Apostolova, I. & Koleva, E. Influence of warming on timberline rising: a case study on *Pinus peuce* Griseb. in Bulgaria. *Phytocoenologia* **30**(3–4), 431–438 (2000).
89. Wardle, P. & Coleman, M. C. Evidence for rising upper limits of four native New Zealand forest trees. *NZ J. Bot.* **30**, 303–314 (1992).
90. Sturm, M., Racine, C. & Tape, K. Increasing shrub abundance in the Arctic. *Nature* **411**, 546–547 (2001).
91. Kennedy, A. D. Antarctic terrestrial ecosystem response to global environmental change. *Annu. Rev. Ecol. Syst.* **26**, 683–704 (1995).
92. Holbrook, S. J., Schmitt, R. J. & Stephens, J. S. Jr Changes in an assemblage of temperate reef fishes associated with a climatic shift. *Ecol. Appl.* **7**, 1299–1310 (1997).
93. Southward, A. J., Hawkins, S. J. & Burrows, M. T. Seventy years' observations of changes in distribution and abundance of zooplankton and intertidal organisms in the western English Channel in relation to rising sea temperature. *J. Therm. Biol.* **20**, 127–155 (1995).
94. Alheit, J. & Hagen, E. Long-term climate forcing of European herring and sardine populations. *Fish. Oceanogr.* **6**(2), 130–139 (1997).
95. Parmesan, C. in *Evolution and Ecology Taking Flight: Butterflies as Model Systems* (eds Boggs, C. L., Watt, W. B. & Ehrlich, P. R.) (Univ. Chicago Press, Chicago, in the press).
96. Thomas, C. D. & Lennon, J. J. Birds extend their ranges northwards. *Nature* **399**, 213 (1999).
97. Hersteinsson, P. & MacDonald, D. W. Interspecific competition and the geographical distribution of red and arctic foxes *Vulpes vulpes* and *Alopex lagopus*. *Oikos* **64**, 505–515 (1992).

Acknowledgements

We thank J. A. Pounds for suggestions and A. Hoppe for assistance with graphics design. Data on spring arrival of birds in northern Germany were provided by O. Huppopp. G.-R.W. received funding from the Swiss NSF and the Swiss Federal Institute of Technology (ETH) Zürich. E.P. acknowledges the financial support of the NSF. A.M. is funded by the 5th FP EU-project POSITIVE and C.P. acknowledges the Centre National de la Recherche Scientifique (CEFE, Montpellier).

Correspondence and requests for materials should be addressed to G.-R.W. (e-mail: walther@geobotanik.uni-hannover.de).

Serotonin_{1A} receptor acts during development to establish normal anxiety-like behaviour in the adult

Cornelius Gross^{*†}, Xiaoxi Zhuang^{*†‡}, Kimberly Stark^{*}, Sylvie Ramboz^{*}, Ronald Oosting^{*}, Lynn Kirby[§], Luca Santarelli^{*}, Sheryl Beck[§] & René Hen^{*}

^{*} Center for Neurobiology and Behavior, Columbia University, New York 10032, USA

[§] Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia 19104, USA

[†] These authors contributed equally to the work

Serotonin is implicated in mood regulation, and drugs acting via the serotonergic system are effective in treating anxiety and depression. Specifically, agonists of the serotonin_{1A} receptor have anxiolytic properties, and knockout mice lacking this receptor show increased anxiety-like behaviour. Here we use a tissue-specific, conditional rescue strategy to show that expression of the serotonin_{1A} receptor primarily in the hippocampus and cortex, but not in the raphe nuclei, is sufficient to rescue the behavioural phenotype of the knockout mice. Furthermore, using the conditional nature of these transgenic mice, we suggest that receptor expression during the early postnatal period, but not in the adult, is necessary for this behavioural rescue. These findings show that postnatal developmental processes help to establish adult anxiety-like behaviour. In addition, the normal role of the serotonin_{1A} receptor during development may be different from its function when this receptor is activated by therapeutic intervention in adulthood.

Agonists of the serotonin_{1A} receptor (5-HT_{1A}) have anxiolytic properties both in humans and animal models^{1,2}, and mice lacking this receptor (5-HT_{1A} knockout) show increased anxiety-like behaviour in a variety of conflict tests^{3–7}. The 5-HT_{1A} is expressed in two distinct neuronal populations in the brain: as an autoreceptor on serotonin-containing neurons of the raphe nuclei, and as a heteroreceptor on non-serotonin-containing neurons of the forebrain (mainly the hippocampus, septum and cortex). Activation of both receptor populations results in membrane hyperpolarization and decreased neuronal excitability. In the case of hippocampal heteroreceptors this effect is mediated by coupling to the GIRK2 potassium channel⁸. Evidence for an anxiolytic role of both receptor populations exists², and as a result it has been difficult to infer which receptor population is responsible for the increased anxiety-like behaviour of 5-HT_{1A} knockout mice. Here we use a tissue-specific and conditional 5-HT_{1A} rescue mouse to dissect the role of these two receptor populations in anxiety-like behaviour. We show that forebrain 5-HT_{1A} is sufficient to rescue the anxious-like phenotype of the knockout mice, arguing that this particular receptor population critically modulates anxiety-like behaviour. In addition, we use the conditional nature of the receptor expression in this mouse to show that receptor expression during the early postnatal period is needed to establish normal anxiety-like behaviour in the adult.

Conditional and tissue-specific receptor expression

To create a mouse in which 5-HT_{1A} expression can be directed to specific tissues in a temporally controlled manner, we designed an inducible knockout allele in which transcription of the receptor is blocked by a stop cassette inserted into the 5' leader sequence of the gene and in which tetO binding sites are introduced upstream of the receptor coding sequence. The tetO binding sites allow the receptor messenger RNA in these otherwise knockout mice to be induced in

the presence of the bacterial transcription factor, tTA. Forebrain-specific expression of the 5-HT_{1A} receptor is achieved by crossing these mice with a transgenic mouse expressing tTA under control of the α -calcium-calmodulin-dependent protein kinase II (α CaMKII) promoter⁹ (Fig. 1a). This double transgenic line, which we call the 'rescue' line, exhibits 5-HT_{1A} binding which roughly mimics the wild-type receptor binding pattern in the hippocampus and cortex, but lacks expression in the raphe nuclei of the brainstem (Fig. 1b). 5-HT_{1A} expression is also seen in the striatum and lateral amygdala, consistent with the known activity of the α CaMKII promoter in these structures. Importantly, the DNA binding function of the tTA protein is blocked by the drug doxycycline¹⁰, so that treatment of the rescue mice with doxycycline (40 mg per kg food for two months) abolishes 5-HT_{1A} expression (Fig. 1b).

To assess whether the receptor in the rescue mice is functional, we examined the responses of CA1 pyramidal neurons to the 5-HT_{1A} full agonist, 5-carboxytryptamine (5-CT), using intracellular recording in hippocampal slices. Cells from wild-type mice show membrane hyperpolarization in response to 5-CT, an effect which is significantly reduced in cells from the knockout mice and partially restored in cells from the rescue mice (Fig. 2a; Kruskal–Wallis analysis of variance (ANOVA), $H_2 = 12.94$, $P < 0.002$; followed by Dunns' post hoc test). Analysis of histograms of the cellular response from the three genotypes reveals a bimodal distribution in the rescue mice where approximately half the cells show a wild-type response to 5-CT and the rest show no response (see Fig. 2b–d); data from the rescue mice failed the Kolmogorov–Smirnov normality test at $P < 0.001$; a χ^2 analysis with $\Sigma_2 = 15.41$ and $P < 0.001$ was significant when the cells were grouped as responders (that is, hyperpolarization > 1) or non-responders (that is, hyperpolarization = 0), indicating that there is a significant difference in the frequency distribution of the data between the cells recorded from the different genotypes. This bimodal response is consistent with the mosaic, or 'patchy', pattern of receptor binding seen in the hippocampus of the rescue mice (Fig. 1b), and argues that those cells showing receptor binding in the rescue mice exhibit a full rescue of the cellular phenotype of the knockout mice.

[‡] Present address: Department of Neurobiology, Pharmacology, and Physiology, University of Chicago, Illinois 60637, USA.

To confirm the absence of functional 5-HT_{1A}R in the raphe nuclei, as suggested by the undetectable receptor binding in the brainstem (Fig. 1b), we examined the responses of raphe neurons to 5-CT using intracellular recording in brainstem slices. Cells from wild-type mice show membrane hyperpolarization in response to 5-CT, an effect which is absent in cells from both knockout and rescue mice (Fig. 2e; ANOVA, $F_{2,12} = 30.6$, $P < 0.0001$, followed by Student–Newman–Keuls post hoc test). To further substantiate the absence of raphe autoreceptors in the rescue mice we also assessed their hypothermic response to the selective 5-HT_{1A}R agonist, 8-OH-DPAT (0.5 mg kg⁻¹, subcutaneously, s.c.). In mice the hypothermic response to this drug has been shown to be mediated by 5-HT_{1A} autoreceptors on serotonergic neurons of the raphe nuclei¹¹. As expected, wild-type mice show a significant drop in body temperature after 8-OH-DPAT administration (Fig. 2f), an effect that is absent in both the knockout and rescue mice (ANOVA, $F_{2,17} = 13.4$, $P < 0.003$).

Forebrain receptor restores normal anxiety-like behaviour

After demonstrating the restoration of functionally active receptors in the forebrain but not in the raphe of the rescue mice, we sought to determine whether this forebrain expression was sufficient to reverse the behavioural phenotype of the knockout animals. We had shown previously that 5-HT_{1A}R knockout mice show increased anxiety-like behaviour in three conflict tests: open field, elevated-plus maze and novelty-suppressed feeding^{4,7}. To examine the anxiety-like behaviour of the rescue mice in these three paradigms, we tested littermates derived from parents heterozygous for the inducible 5-HT_{1A}R allele, one parent of which is also heterozygous for the α CaMKII-tTA transgene. In each of the three conflict tests

the rescue mice show wild-type levels of anxiety-like behaviour, demonstrating a reversal of the anxiety-like phenotype of the knockout mice. In the open field, 5-HT_{1A}R knockout mice move less than wild-type mice in the aversive centre of the open arena (Fig. 3a; ANOVA, $F_{2,45} = 4.34$, $P = 0.019$). In addition, the knockout mice show an overall decrease in locomotion in the open field (Fig. 3b; ANOVA, $F_{2,45} = 6.96$, $P = 0.0023$), an effect which is likely to reflect their response to novelty, as their locomotion in the home cage is unaltered⁴. Rescue mice show wild-type levels of locomotion in the centre of the open field as well as wild-type levels of total locomotion (Fig. 3a, b). In the elevated-plus maze knockout mice make fewer entries into the aversive open arms, whereas the behaviour of the rescue mice is indistinguishable from that of wild-type animals (Fig. 3c; ANOVA, $F_{2,45} = 4.2$, $P = 0.021$). Total entries into the closed and open arms did not differ between the genotypes (ANOVA, $F_{2,45} = 0.43$, $P = 0.44$). Similarly, in the novelty-suppressed feeding test, knockout mice take significantly longer to begin eating in a novel environment, whereas rescue mice show a latency to feed similar to that of wild-type animals (Fig. 3d; ANOVA, $F_{2,21} = 4.5$, $P = 0.024$). Food consumption in the home cage was measured for 5 min after exposure to the novel environment and did not differ between genotypes (ANOVA, $F_{2,21} = 1.8$, $P = 0.19$). Together, these data show that 5-HT_{1A}R in the forebrain is sufficient to restore normal anxiety-like behaviour in the knockout background, and suggest that forebrain receptors play a critical role in modulating anxiety-like behaviour in wild-type mice. In addition, the ability of mosaic receptor expression to rescue the behavioural phenotype of the knockout mice argues either for a cell non-autonomous role of the receptor, or that a partial restoration of brain function is sufficient for the rescue.

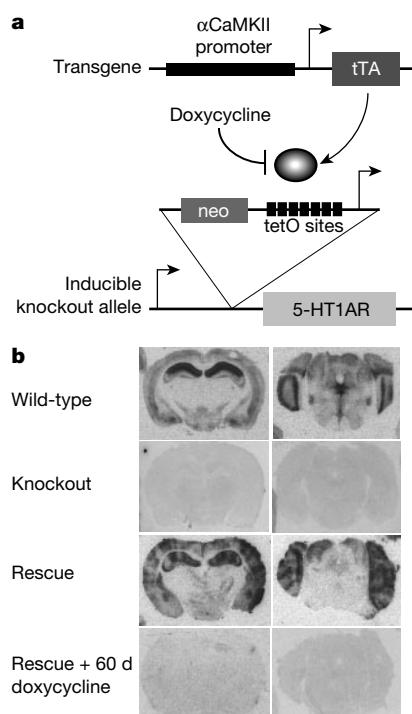


Figure 1 Genetic rescue strategy and 5-HT_{1A}R expression. **a**, A mouse carrying a rescuable knockout allele of the 5-HT_{1A}R was crossed with a transgenic mouse carrying the bacterial tTA protein under the control of the α CaMKII promoter to achieve expression of the 5-HT_{1A}R in a tissue-specific and temporally controlled manner. **b**, Expression of 5-HT_{1A}R in the rescue mice. ¹²⁵I-MPPI receptor autoradiography²⁸ reveals high levels of receptor expression in the hippocampus and cortex of the rescue mice. In contrast, the receptor is undetectable in the raphe nuclei of the rescue mice. Treatment of the rescue mice with doxycycline for two months eliminates receptor binding.

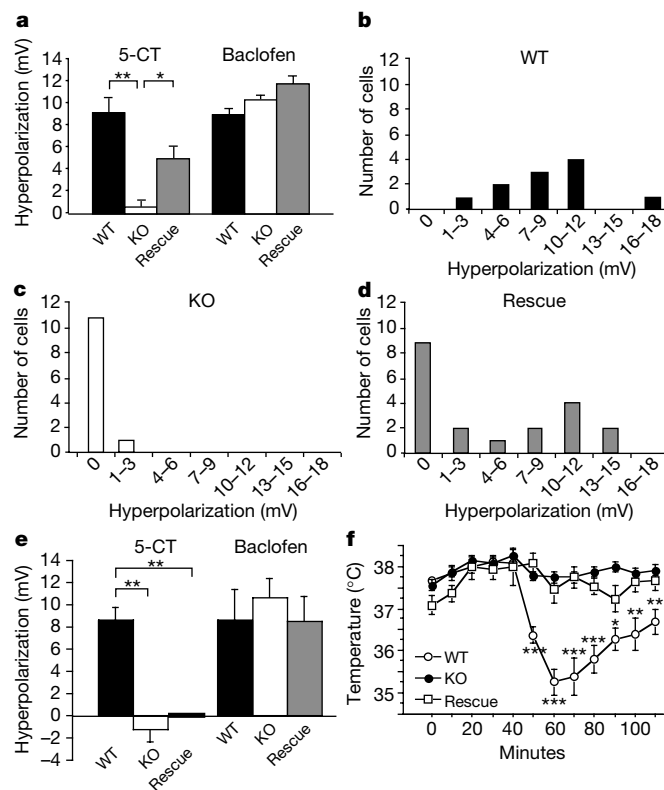


Figure 2 Functional 5-HT_{1A}R in the hippocampus, but not in raphe nuclei of rescue mice. **a**, Hyperpolarization response to 5-CT and baclofen in CA1 hippocampal pyramidal cells of wild-type, knockout and rescue mice. **b–d**, Histograms of the data summarized in **a** showing the distribution of 5-CT responses. **e**, Hyperpolarization response to 5-CT and baclofen in dorsal raphe cells. **f**, Hypothermic response to 8-OH-DPAT. Asterisks, see Methods.

Adult receptor not required to maintain rescue phenotype

Next, we used the conditional nature of the receptor in the rescue mice to assess the contribution of 5-HT_{1A} during adulthood and development to the rescue phenotype. First, we examined the contribution of the adult receptor by treating 10–12-week-old mice of all three genotypes for two months with doxycycline (40 mg per kg food) to shut off receptor expression (Fig. 4a, see also Fig. 1b). Receptor autoradiography at different time points after the start of doxycycline treatment shows that receptor binding disappears with a half-life, $t_{1/2}$, of approximately 15 days (data not shown). We note that when these mice are tested in the open field, elevated-plus maze and novelty-suppressed feeding tests, the rescue mice continue to show wild-type levels of anxiety-like behaviour, demonstrating that the receptor in the adult is not required to maintain the rescue phenotype (see Fig. 4b–e): (1) open field, per cent path length in the centre, ANOVA, $F_{2,45} = 3.57$, $P = 0.037$ and total path length, ANOVA, $F_{2,45} = 11.35$, $P = 0.0001$; (2) elevated-plus maze, per cent entries in open arms, ANOVA, $F_{2,45} = 4.81$, $P = 0.013$ and total entries, ANOVA, $F_{2,45} = 0.51$, $P = 0.60$; (3) novelty-suppressed feeding, latency, ANOVA, $F_{2,45} = 8.78$, $P = 0.0006$ and home cage food consumption, ANOVA, $F_{2,45} = 0.93$, $P = 0.40$. This result is consistent with the inability of the 5-HT_{1A} antagonist, WAY100635 (0.1 mg kg^{-1} , s.c.), to alter anxiety-like behaviour in the open field, elevated-plus maze and novelty-suppressed feeding tests in wild-type mice of this strain (see Supplementary Information), and is in line with the lack of anxiogenic effects of 5-HT_{1A} antagonists reported by other researchers^{12,13}. Together, these results lead us to hypothesize that in the adult animals, activation of forebrain 5-HT_{1A} receptors by endogenous serotonin may not affect anxiety-like behaviour. We emphasize, however, that this conclusion does not rule out the possibility that activation of the receptor by exogenous treatment with 5-HT_{1A} agonists may still be anxiolytic. Indeed, we have shown that wild-type mice show decreased anxiety-like behaviour in the novelty-suppressed feeding test after chronic, but not acute, treatment with the 5-HT_{1A} agonist, 8-OH-DPAT (see Supplementary Information).

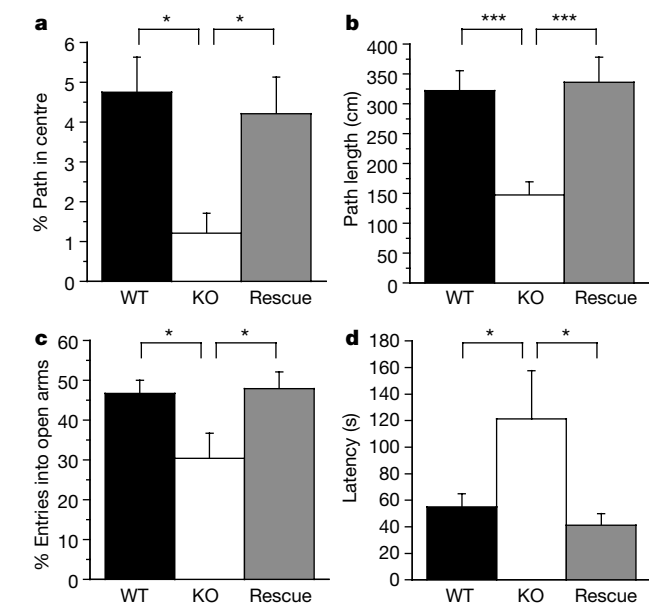


Figure 3 Anxiety-like behaviour in the rescue mice is normal. **a**, Per cent path length in the centre of the open field. **b**, Total path length in the open field. **c**, Per cent entries into the open arms of the elevated-plus maze. **d**, Latency to feed in the novelty-suppressed feeding test. Asterisks, see Methods.

Receptor required during development

The fact that removing or blocking 5-HT_{1A} function does not mimic the knockout phenotype suggests that the receptor functions earlier in development to establish normal adult anxiety-like behaviour. Turning off receptor expression during this critical developmental period should, therefore, reverse the rescue phenotype. To test this hypothesis, receptor expression was turned off during the embryonic and early postnatal period by treating breeding pairs with doxycycline and switching their pups to normal food after weaning at postnatal day 21 (P21, Fig. 5a). The absence of receptor in rescue mice treated with doxycycline through the mother was confirmed at postnatal day 14 (data not shown). Receptor autoradiography at different time points after removing doxycycline shows that the 5-HT_{1A} appears with a $t_{1/2}$ of approximately 4 days (data not shown). When tested as adults in the open field, elevated-plus maze, and novelty-suppressed feeding test, these rescue mice behave indistinguishably from knockout mice (Fig. 5b–e): (1) open field, per cent path length in the centre, ANOVA, $F_{2,56} = 11.9$, $P < 0.0001$ and total path length, ANOVA, $F_{2,56} = 3.37$, $P = 0.042$; (2) elevated-plus maze, per cent entries in open arms, ANOVA, $F_{2,55} = 4.45$, $P = 0.016$; (3) novelty-suppressed feeding test, latency, ANOVA, $F_{2,40} = 6.07$, $P = 0.0050$ and home cage food consumption, ANOVA, $F_{2,40} = 2.53$, $P = 0.093$. These results demonstrate that receptor expression before P21 is required for successful rescue of the knockout phenotype.

Expression of the 5-HT_{1A} in the forebrain of wild-type mice first appears at embryonic day 17 (E17), after which expression gradually increases to adult levels during the postnatal period (Fig. 6). Receptor expression in the rescue mice, on the other hand, first becomes detectable in the hippocampus and cortex at P5, and reaches significant levels at P15, consistent with the time

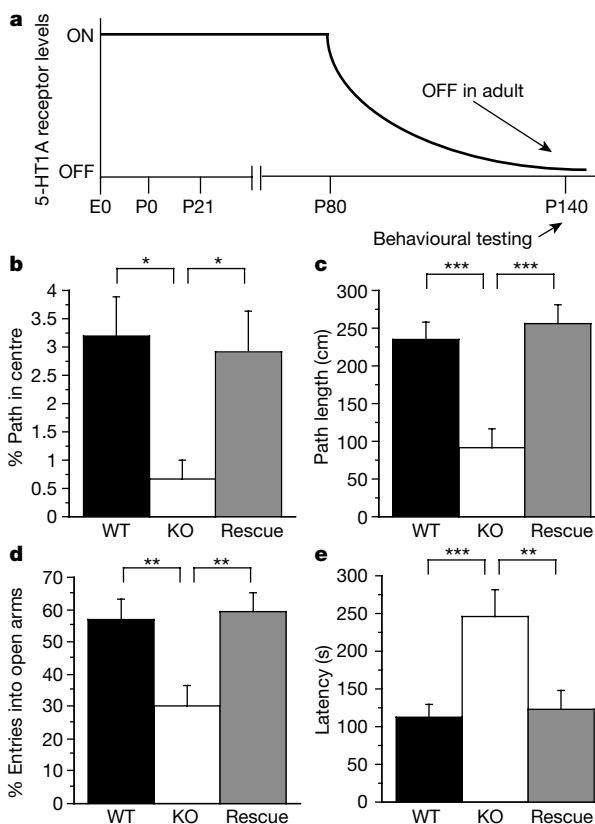


Figure 4 Turning off 5-HT_{1A} in the adult does not alter anxiety-like behaviour. **a**, Adult wild-type, knockout and rescue mice were treated with doxycycline for two months and submitted to behavioural testing in: **b–c**, the open field; **d**, the elevated-plus maze; **e**, the novelty-suppressed feeding test. Asterisks, see Methods.

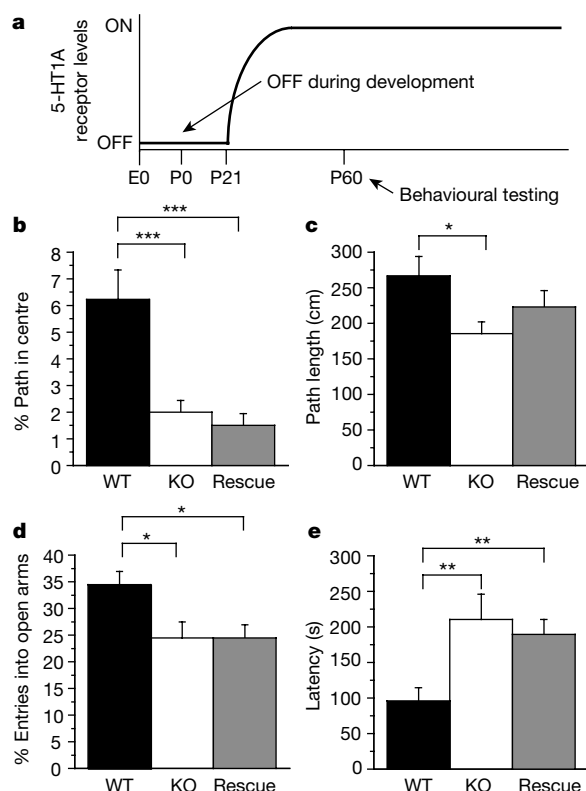


Figure 5 Turning off 5-HT1AR during development is sufficient to reverse the rescue phenotype. **a**, To turn off the receptor during the embryonic and early postnatal period, breeding pairs were fed doxycycline and offspring were weaned, switched to normal food at postnatal day 21 (P21), and tested in adulthood in: **b–c**, the open field; **d**, the elevated-plus maze; **e**, the novelty-suppressed feeding test. Asterisks, see Methods.

course of appearance of the α CaMKII gene¹⁴. These results suggest that the critical period for establishment of the rescue phenotype is between P5 and P21.

Discussion

The idea that the early postnatal period is a critical time for the establishment of lifelong anxiety behaviour has considerable support. In humans, some evidence suggests that early life stressors increase susceptibility to anxiety and mood disorders in adulthood^{15,16}. In monkeys, increased variability in foraging demand on the mother leads to increased anxiety-like behaviour in adult offspring¹⁷. Moreover, in rats, increased anxiety-like behaviour is seen in the adult offspring of mothers who lick and groom their pups less frequently¹⁸. We note that serotonin receptors are implicated in at least part of the effect of maternal licking and grooming in rats¹⁹, suggesting that serotonin might have an important role in establishing proper neural circuits during this period. Mature serotonergic innervation of the cortex and hippocampus occurs during the first

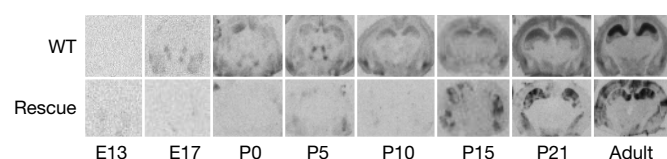


Figure 6 Expression of the 5-HT1AR during development by ¹²⁵I-MPPI receptor autoradiography²⁴. In wild-type mice, the receptor is first detected in the forebrain at embryonic day 17 (E17) after which levels gradually increase to near adult levels at postnatal day 21 (P21). In rescue mice, the receptor is first detected in the forebrain at P5, and reaches significant levels at P15.

two postnatal weeks²⁰, a process that is accompanied by a twofold increase in serotonin^{19,21}. Insufficient or excessive levels of serotonin during the postnatal period lead to long-lasting changes in brain plasticity, including the disruption of sensory maps in the visual and sensory cortices^{22,23}, decreased dendritic spine density²⁴ and blunted short-term plasticity (paired pulse facilitation) in the dentate gyrus²⁵, and a reduction in the number of calretinin-positive cortical neurons with bipolar morphology²⁶. Decreased dendritic spine density in the dentate gyrus has also been reported after postnatal treatment with 5-HT1AR antagonists²⁷, suggesting that the 5-HT1AR may be a critical mediator of at least some of the developmental effects of serotonin.

Our results suggest that normal anxiety-like behaviour in the adult requires proper signalling by serotonin via forebrain 5-HT1A receptors during the early postnatal period. Presumably, stimulation of the receptor during this period is essential to set in motion a cascade of events which leads to long-lasting changes in brain chemistry or structure that are essential for normal anxiety-like behaviour throughout life. We note that in humans chronic treatment with direct 5-HT1AR agonists such as buspirone or indirect agonists such as fluoxetine is required for therapeutic efficacy, raising the possibility that both during development and in adulthood activation of 5-HT1A receptors can elicit long-term plastic changes that decrease anxiety-like behaviour. □

Methods

Electrophysiology

Adult mice (>ten weeks of age) were decapitated, the brain removed and placed in ice-cold buffer containing 124 mM NaCl, 3 mM KCl, 1.25 mM NaH₂PO₄, 2 mM MgSO₄, 2.5 mM CaCl₂, 10 mM dextrose and 26 mM NaHCO₃. Raphe and hippocampal coronal slices (500 μ m) were cut on a vibratome and placed in a holding chamber until used. After at least one hour of equilibration a slice was transferred to a recording chamber where it was constantly perfused with buffer bubbled with 95% O₂/5% CO₂ at a flow rate of 1.5 ml min⁻¹. Electrodes were pulled from borosilicate capillary tubing and filled with 2 M KCl to obtain resistances of 60–140 M Ω . Area CA1 hippocampal pyramidal or dorsal raphe cells were impaled and hyperpolarized to facilitate sealing of the cell. Electrical signals were amplified and the data set to a computer for online recording and analysis. 5-HT1A and GABA_B receptor-mediated hyperpolarizations were measured following bath application of 5-CT (100 nM) or baclofen (10 μ M) using pCLAMP software (Axon Instruments). For measurement of hypothermia in response to 8-OH-DPAT, rectal temperature was measured in singly housed mice every 10 min and the drug was administered to all animals at $t = 40$ min. Doses of 5-CT, baclofen and 8-OH-DPAT were shown to elicit maximal responses.

Behavioural testing

We tested 8–10-week-old male and female mice (18–20 weeks for rescue mice treated during adulthood with doxycycline; all mice of 129/Sv; C57BL/6J; CBA/J strain) homozygous for either the wild-type (WT) or inducible knockout (KO) 5-HT1AR alleles with or without a single copy of the α CaMKII-tTA transgene (WT, WT + α CaMKII-tTA, KO, KO + α CaMKII-tTA = rescue). No significant sex differences were observed. Data from WT and WT + α CaMKII-tTA mice were combined because these two genotypes were indistinguishable in all behavioural tests. Mice were tested sequentially in the open field (60 min in a 40 cm $\times$ 40 cm white arena coupled to an automated video tracking system; data are expressed as the average value of 12 five-minute bins), elevated-plus maze⁴ (5 min), and novelty-suppressed feeding test⁷ at one-week intervals. The novelty-suppressed feeding test consists of a 40 cm $\times$ 60 cm open arena with a food pellet placed in the centre. Food-deprived mice (24 h) are placed in the arena and the latency to begin chewing food is recorded. Chewing is scored when the mouse is sitting on its haunches and biting with the use of forepaws. Immediately after beginning to eat, mice are transferred to their home cage and the amount of food consumed in 5 min is measured (home cage food consumption). Statistical significance was assessed by analysis of variance (ANOVA), followed by Fisher's post hoc test, unless otherwise indicated, with * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Received 23 October 2001; accepted 22 January 2002.

- Feighner, J. P. & Boyer, W. F. Serotonin-1A anxiolytics: an overview. *Psychopathology* **22** (Suppl. 1), 21–26 (1989).
- Menard, J. & Treit, D. Effects of centrally administered anxiolytic compounds in animal models of anxiety. *Neurosci. Biobehav. Rev.* **23**, 591–613 (1999).
- Parks, C. L., Robinson, P. S., Sibille, E., Shen, T. & Toth, M. Increased anxiety of mice lacking the serotonin1A receptor. *Proc. Natl Acad. Sci. USA* **95**, 10734–10739 (1998).
- Ramboz, S. et al. Serotonin receptor 1A knockout: an animal model of anxiety-related disorder. *Proc. Natl Acad. Sci. USA* **95**, 14476–14481 (1998).
- Heisler, L. K. et al. Elevated anxiety and antidepressant-like responses in serotonin 5-HT1A receptor mutant mice. *Proc. Natl Acad. Sci. USA* **95**, 15049–15054 (1998).
- Sibille, E., Pavlidis, C., Benke, D. & Toth, M. Genetic inactivation of the Serotonin(1A) receptor in

- mice results in downregulation of major GABA(A) receptor alpha subunits, reduction of GABA(A) receptor binding, and benzodiazepine-resistant anxiety. *J. Neurosci.* **20**, 2758–2765 (2000).
7. Gross, C., Santarelli, L., Brunner, D., Zhuang, X. & Hen, R. Altered fear circuits in 5-HT1A receptor KO mice. *Biol. Psychiatry* **48**, 1157–1163 (2000).
 8. Luscher, C., Jan, L. Y., Stoffel, M., Malenka, R. C. & Nicoll, R. A. G protein-coupled inwardly rectifying K⁺ channels (GIRKs) mediate postsynaptic but not presynaptic transmitter actions in hippocampal neurons. *Neuron* **19**, 687–695 (1997).
 9. Mayford, M. *et al.* Control of memory formation through regulated expression of a CaMKII transgene. *Science* **274**, 1678–1683 (1996).
 10. Kistner, A. *et al.* Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 10933–10938 (1996).
 11. Martin, K. F., Phillips, I., Hearson, M., Prow, M. R. & Heal, D. J. Characterization of 8-OH-DPAT-induced hypothermia in mice as a 5-HT1A autoreceptor response and its evaluation as a model to selectively identify antidepressants. *Br. J. Pharmacol.* **107**, 15–21 (1992).
 12. Fletcher, A. *et al.* Electrophysiological, biochemical, neurohormonal and behavioural studies with WAY-100635, a potent, selective and silent 5-HT1A receptor antagonist. *Behav. Brain Res.* **73**, 337–353 (1996).
 13. Cao, B. J. & Rodgers, R. J. Influence of 5-HT1A receptor antagonism on plus-maze behaviour in mice. II. WAY 100635, SDZ 216-525 and NAN-190. *Pharmacol. Biochem. Behav.* **58**, 593–603 (1997).
 14. Bayer, K. U., Lohler, J., Schulman, H. & Harbers, K. Developmental expression of the CaM kinase II isoforms: ubiquitous gamma- and delta-CaM kinase II are the early isoforms and most abundant in the developing nervous system. *Brain Res. Mol. Brain Res.* **70**, 147–154 (1999).
 15. Tweed, J. L., Schoenbach, V. J., George, L. K. & Blazer, D. G. The effects of childhood parental death and divorce on six-month history of anxiety disorders. *Br. J. Psychiatry* **154**, 823–828 (1989).
 16. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. Childhood parental loss and adult psychopathology in women. A twin study perspective. *Arch. Gen. Psychiatry* **49**, 109–116 (1992).
 17. Coplan, J. D., Rosenblum, L. A. & Gorman, J. M. Primate models of anxiety. Longitudinal perspectives. *Psychiatric Clin. North. Am.* **18**, 727–743 (1995).
 18. Caldji, C. *et al.* Maternal care during infancy regulates the development of neural systems mediating the expression of fearfulness in the rat. *Proc. Natl Acad. Sci. USA* **95**, 5335–5340 (1998).
 19. Mitchell, J. B., Iny, L. J. & Meaney, M. J. The role of serotonin in the development and environmental regulation of type II corticosteroid receptor binding in rat hippocampus. *Brain Res. Dev. Brain Res.* **55**, 231–235 (1990).
 20. Lidov, H. G. & Molliver, M. E. An immunohistochemical study of serotonin neuron development in the rat: ascending pathways and terminal fields. *Brain Res. Bull.* **8**, 389–430 (1982).
 21. Hohmann, C. F., Hamon, R., Batshaw, M. L. & Coyle, J. T. Transient postnatal elevation of serotonin levels in mouse neocortex. *Brain Res.* **471**, 163–166 (1988).
 22. Cases, O. *et al.* Lack of barrels in the somatosensory cortex of monoamine oxidase A-deficient mice: role of a serotonin excess during the critical period. *Neuron* **16**, 297–307 (1996).
 23. Upton, A. L. *et al.* Excess of serotonin (5-HT) alters the segregation of ipsilateral and contralateral retinal projections in monoamine oxidase A knock-out mice: possible role of 5-HT uptake in retinal ganglion cells during development. *J. Neurosci.* **19**, 7007–7024 (1999).
 24. Yan, W., Wilson, C. C. & Haring, J. H. Effects of neonatal serotonin depletion on the development of rat dentate granule cells. *Brain Res. Dev. Brain Res.* **98**, 177–184 (1997).
 25. Haring, J. H. & Yan, W. Dentate granule cell function after neonatal treatment with parachloroamphetamine or 5,7-dihydroxytryptamine. *Brain Res. Dev. Brain Res.* **114**, 269–272 (1999).
 26. Durig, J. & Hornung, J. P. Neonatal serotonin depletion affects developing and mature mouse cortical neurons. *Neuroreport* **11**, 833–837 (2000).
 27. Yan, W., Wilson, C. C. & Haring, J. H. 5-HT1a receptors mediate the neurotrophic effect of serotonin on developing dentate granule cells. *Brain Res. Dev. Brain Res.* **98**, 185–190 (1997).
 28. Kung, M. P., Frederick, D., Mu, M., Zhuang, Z. P. & Kung, H. F. 4-(2'-methoxy-phenyl)-1-[2'-(n-2"-pyridinyl)-p-idobenzamido]-ethyl-piperazine ([125I]p-MPPI) as a new selective radioligand of serotonin-1A sites in rat brain: *in vitro* binding and autoradiographic studies. *J. Pharmacol. Exp. Ther.* **272**, 429–437 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank I. Mendez for her help with mouse husbandry and behavioural analysis. This research was supported by grants from NIMH (C.G., S.B., R.H.), NINDS (S.B.), NIDA (R.H.), and NARSAD (X.Z., L.S., R.H.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.H. (e-mail: rh95@columbia.edu).

Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues

Max P. Bernstein^{*†}, Jason P. Dworkin^{*†}, Scott A. Sandford[†], George W. Cooper[†] & Louis J. Allamandola[†]

^{*} The Center for the Study of Life in the Universe, SETI Institute, 2035 Landings Drive, Mountain View, California 94043, USA

[†] NASA-Ames Research Center, Mail Stop 245-6, Moffett Field, California 94035-1000, USA

The delivery of extraterrestrial organic molecules to Earth by meteorites may have been important for the origin and early evolution of life¹. Indigenous amino acids have been found in meteorites²—over 70 in the Murchison meteorite alone³. Although it has been generally accepted that the meteoritic amino acids formed in liquid water⁴ on a parent body, the water in the Murchison meteorite is depleted in deuterium⁵ relative to the indigenous organic acids^{6,7}. Moreover, the meteoritic evidence⁸ for an excess of laevo-rotatory amino acids is hard to understand in the context of liquid-water reactions on meteorite parent bodies. Here we report a laboratory demonstration that glycine, alanine and serine naturally form from ultraviolet photolysis of the analogues of icy interstellar grains. Such amino acids would naturally have a deuterium excess similar to that seen in interstellar molecular clouds, and the formation process could also result in enantiomeric excesses if the incident radiation is circularly polarized. These results suggest that at least some meteoritic amino acids are the result of interstellar photochemistry, rather than formation in liquid water on an early Solar System body.

As the most ancient and pristine bulk material studied in the laboratory, primitive meteorites are the clearest windows to the birth of the Solar System. Planetary systems such as our own are believed to form from the collapse of an interstellar dense molecular cloud composed of gas and sub-micrometre sized grains. In such 'dark' clouds the temperatures are low ($T < 50$ K), and all but the

most volatile species (that is, H_2 , He, Ne) condense onto grains, coating them with a thin layer of ice⁹. This ice is composed primarily of amorphous H_2O , but usually also contains a variety of other simple molecules^{9,10}, such as CO_2 , CO, CH_3OH , and NH_3 . Laboratory studies¹¹ and astronomical observations^{10,12} indicate that radiation processing of such ices can create complex organic compounds¹³. Many of the organic molecules that are present in carbonaceous chondrites (primitive carbon-rich meteorites) and comet and asteroid dust are thought to come, at least in part, from the ice and complex compounds constructed in the interstellar medium (ISM).

Perhaps the most convincing molecular evidence for the interstellar heritage of meteoritic molecules is their high deuterium (D) enrichment^{3,14}. At the low temperatures in dense molecular clouds deuterium fractionation is efficient and elevated D/H ratios are seen in grain mantles¹⁵; such increased ratios are also found in several gas-phase interstellar molecules, including amino acid precursors, such as formaldehyde and ammonia¹⁶. Although it had been accepted that the deuterium in meteoritic organics indicated that their precursors formed in the ISM, the actual hydroxy and amino acids are still commonly believed to have formed on the asteroid or comet parent body from reactions in liquid water⁴ that, at least in Murchison, was apparently deuterium poor⁵. It is difficult to explain how these compounds retain relatively high amounts of deuterium, let alone how it is distributed. For example, it seems contradictory that the hydroxy acids in the Murchison meteorite have one-third as much deuterium as the amino acids, and yet have a lower rate of deuterium exchange in water than amino acids^{3,7}. If, however, the hydroxy and amino acids had formed in the ISM their deuterium enrichment would be a logical consequence of the photochemistry of already deuterium-enriched pre-solar ices.

The laboratory experiments described here were designed to elucidate potential pathways from interstellar chemistry to the organic molecules extracted from meteorites. We have conducted laboratory experiments at temperatures, pressures and radiation conditions that are representative of the interstellar clouds from which planetary systems form. In a series of experiments, gases were vapour deposited onto a substrate at 15 K forming an ice film consisting primarily of amorphous H_2O with other compounds over a range of concentrations (0.5–5% NH_3 , 5–10% CH_3OH and 0.5–5% HCN, relative to H_2O). These solid mixtures are

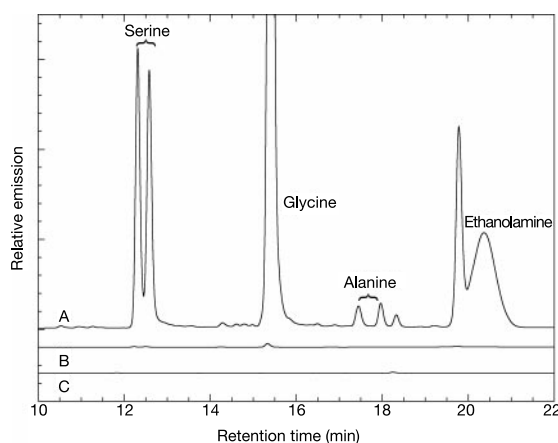


Figure 1 Amino acids are formed by the UV photolysis of a realistic interstellar ice analogue. This is demonstrated by the comparison of HPLC traces of derivatized amines resulting from: trace A, the UV photolysis of an $H_2O:CH_3OH:NH_3:HCN = 20:2:1:1$ ice showing that the amino acids serine, glycine and alanine, as well as other molecules, are produced; trace B, a control of the same ice with no UV photolysis; and trace C, a procedural blank. In trace A, a single peak indicates the presence of the amino acid

glycine but the chiral fluorescent tag, which separates enantiomeric amines, causes the racemic serine and alanine to appear as pairs of peaks. Differing molar absorptivities of the labelled D,L serine and alanine diastereomers account for asymmetry in the peak pairs. The unlabelled peaks are unidentified amines. The racemic nature of the serine and alanine and the absence of prominent peaks in traces B and C indicate that contamination is not significant.

representative of the composition of interstellar ice mantles in dense clouds and towards forming stars (protostars). For example, relative to H_2O , NH_3 has been observed at 10–30% in the ice around the massive protostars NGC7538¹⁷ and GCS3¹⁸, and CH_3OH has been commonly observed in clouds around forming stars¹⁹. Thus, H_2O , CH_3OH and NH_3 are reasonable starting materials because they are among the most abundant molecules frozen onto grains in the dense ISM. In addition, it is reasonable to include HCN because it is abundant in cometary coma and the dense ISM²⁰, where the majority of HCN should be frozen onto grains. Other simple molecules present in interstellar ices at comparable or greater abundances (CO , CO_2) rapidly form *in situ* in our experiments as a result of ultraviolet (UV) photolysis^{13,21}.

The interstellar ice simulations were performed in a cryogenic sample chamber, and great care was taken to exclude contamination and ensure that the starting materials did not react prior to deposition. Amorphous H_2O ices containing CH_3OH , HCN and NH_3 were vapour-deposited at temperatures and pressures representative of dense molecular clouds (15 K and 10^{-8} torr, respectively) while simultaneously exposed to UV radiation characteristic of the ISM. On warming under dynamic vacuum, the ice sublimates, leaving behind an organic residue that was analysed by gas chromatography–mass spectrometry (GC-MS) and high-precision liquid chromatography (HPLC). In GC-MS, the identity of molecules are ascertained from both their retention time and mass fragmentation patterns. The HPLC allows us to identify amines bearing a fluorescent label by the retention time with the co-injection of authentic standards.

N-formyl glycine, cycloserine (4-amino-3-isoxazolidinone) and glycerol were detected in the organic residue before hydrolysis, and glycine, alanine, serine, glycerol, ethanolamine and glyceric acid were observed by GC-MS after hydrolysis. Amino malonic acid and hydantoins were also tentatively identified. After hydrolysis, glycine, alanine, serine urea and ethanolamine were detected by HPLC (Fig. 1). The carbon yield of glycine, the most abundant amino acid when starting from methanol, was typically about 0.5%. Most biological amino acids are chiral, meaning that they exist in two mirror image (enantiomeric) forms—dextro-rotatory (D) and laevo-rotatory (L). Living systems employ L-amino acids almost exclusively, while the amino acid products of (abiotic) laboratory syntheses are usually a 50:50 mixture of D and L (a racemic mixture). The chiral amino acids produced in our experiments—serine and alanine—were determined to be racemic, $(100.1 \pm 1.6)\%$, within the integration error, providing evidence that they were not contaminants. This is also demonstrated by controls showing that if the UV was omitted the amino acids detected by HPLC (Fig. 1, trace B) are 1,000 times less abundant than in otherwise identical experiments where UV light was included during deposition. In addition, GC-MS of organics produced in experiments consisting of fully ^{13}C labelled starting material showed fully labelled products, further confirming that contamination is minimal.

We detect N-formyl glycine and cycloserine before hydrolysis, but observe that the abundances of these molecules diminish with hydrolysis along with a concomitant appearance of glycine and serine. This strongly suggests that at least part of the amino acids we observe by HPLC after hydrolysis had formed as individual molecular species, and are not degradation products of macromolecules, such as HCN polymer²². We recently reported rates of amino acid decomposition on exposure to vacuum UV under interstellar conditions²³. This allowed us to put constraints on the lifetimes of amino acids in the ISM, both in the gas-phase and in ices thin enough to be penetrated by UV light. Those results are consistent with this report suggesting that meteoritic amino acids could have formed in the ISM. However, taken together they suggest that the precursors are generated by UV radiation in the ice at low temperature, but the amino acids themselves probably do not form until the ice is warmed, and photo-products can react with one another. As

our ices contain HCN, NH_3 and H_2CO (formaldehyde, from the photolysis of CH_3OH), it is possible that the formation mechanism is related to the Strecker synthesis—the reaction of these three molecules in liquid water to make glycine⁴. However, mechanisms involving radicals have also been proposed²⁴.

Recent careful measurements have established that certain indigenous, non-biological, meteoritic amino acids favour the L form⁸. This would be very difficult to explain using exclusively aqueous parent-body reactions, but astronomical observations of circularly polarized radiation in OMC-1²⁵ have lent credence to the notion that a radiative process might account for such enantiomeric excesses of amino acids in meteorites²⁶. However, only inefficient mechanisms based on enantio-selective photodestruction of a racemic starting population have been demonstrated in the laboratory²⁷. Thus, it had been generally assumed that a radiative process was synthesizing such chiral molecules⁸, but formation of racemic amino acids under simulated interstellar conditions in the laboratory had not been demonstrated until now.

Thus, ice photochemistry potentially provides a single simple explanation for the presence, deuterium-enrichment, and enantiomeric excess of at least some amino acids and related compounds in meteorites. Furthermore, whereas parent body reactions in liquid water require conditions that are relatively uncommon, dense molecular clouds, being thousands of light years across, are vast, ubiquitous, chemical reactors. As the materials from which all stellar systems are made pass through such clouds, amino acids should have been incorporated into all other planetary systems, and thus been available for the origin of life. □

Methods

Materials

The water was purified with a Millipore Milli-Q water system to 18.2 M Ω . Methanol (Aldrich, HPLC grade, 99.93%) and water were freeze-pump-thawed three times before being transferred under vacuum into the glass bulb holding the matrix gases. Anhydrous NH_3 (Matheson) was transferred under vacuum into the glass bulb holding the sample gases without any purification. HCN was generated from the addition of concentrated H_2SO_4 (Poly Research Corp.) to KCN (Mallinckrodt) under vacuum followed by vacuum distillation of the HCN. A 0.03 mm Ni (99.9%) foil (Goodfellow) was used as a substrate.

Sample preparation

Our interstellar ice simulation chamber is a high-vacuum matrix isolation apparatus described in detail elsewhere²¹. Two gas bulbs with H_2O and CH_3OH —one also containing HCN and the other NH_3 —were vapour-deposited simultaneously but separately onto a high purity Ni substrate held at 15 K. Care was taken to ensure that the two different gas mixtures (one with HCN, the other NH_3) did not come into contact with one another until frozen onto the substrate. Infrared spectra of ices deposited onto a CsI substrate demonstrated that they had not reacted during the course of deposition.

The mixed molecular ices were UV-photolysed during deposition with a microwave-powered, flowing hydrogen, discharge lamp²⁸. This lamp produces $\sim 2 \times 10^{15}$ photons $\text{cm}^{-2} \text{s}^{-1}$, the flux being nearly evenly divided between the Lyman α line and a roughly 20-nm-wide molecular transition centred at 160 nm. Under such conditions, the time that elapses between photons arriving in the same molecular neighbourhood is ~ 13 orders of magnitude longer than molecular relaxation times, so multi-photon processes are not relevant. Typically, each sample was photolysed for the equivalent of ~ 30 min ($\sim 4 \times 10^{18}$ photons cm^{-2}) per 0.1 μm of ice. This is a reasonable interstellar dose, corresponding to ~ 500 yr at the edge of a dense cloud and $\sim 1 \times 10^5$ yr at an optical depth of 5 within a dense molecular cloud²⁸.

After deposition and photolysis were complete, the ices were warmed at $\sim 2 \text{ K min}^{-1}$ under dynamic vacuum at $\sim 10^{-8}$ torr to room temperature. Under these experimental conditions the ice sublimates during warm-up, leaving behind an organic residue. At no point during this procedure does the ice melt, nor was the residual organic material exposed to any liquid until the hydrolysis for HPLC analysis. This allowed us to analyse the suite of organic compounds present both before and after the hydrolysis step. The Ni substrate on which the experiment has been performed was then removed from the vacuum system. To guard against contamination of the residue by biological amino acids and other organics, the Ni substrate, all glassware, and tools coming in contact with it, were cleaned of organic material by baking at 550 °C in air for more than 15 h prior to use.

Hydrolysis and residue analysis

The Ni substrate was removed from the vacuum system and cut in two. One half was treated with bis-trimethylsilyltrifluoroacetamide containing 1% trimethyl chlorosilane (Alltech) and anhydrous pyridine in a 1:3 ratio. This solution was sonicated and stirred for 30 min, after which 1 μl was injected into the GC-MS. Such samples were never exposed to liquid water at any time during formation or analysis.

The other half of the Ni substrate was shaken with 18 M Ω Millipore water, then

removed, and the solution added to an equal volume of 12 M HCl (Ultrex II ultrapure reagent, Baker) and heated for 16 h at 100 °C in an organic-free sealed glass ampoule. This material was dried *in vacuo* to remove water, HCl and any volatile organic acids, then analysed by GC-MS as above. GC-MS was performed on a Thermoquest-Finnigan GCQ with an injector temperature of 240 °C and a DB-17ms-60m (J&W Scientific) column at an initial temperature of 70 °C increasing at 5° per min to 240 °C.

Samples for HPLC were hydrolysed as above, and after drying, 100 µl of 10 mM pH 9.5 sodium borate was added to the tube which, after mixing, was re-dried *in vacuo* to remove volatile amines. The sample was analysed via the OPA/NAC fluorescent labelling and chromatography protocols of ref. 30 on a Hewlett Packard 1100 series HPLC with a Supelco Discovery C-18, 5 µm resin, 4.6 × 250 mm analytical column with a 5 µl sample loop. This chiral label reacts with chiral amines (such as alanine and serine), forming diastereomers which can be separated by HPLC. However, differing absorptivities of these labelled diastereomers results in pairs of peaks where the areas of diastereomers are different even though there are equal amounts of the D and L forms.

After a mild acid hydrolysis with the acid concentration diminished by a factor of 100 (0.06 M HCl), glycine and serine were measured at levels 3–4 times lower than for the standard hydrolysis procedure. Even hot water was adequate to release glycine, albeit an order of magnitude less than the standard hydrolysis.

Received 29 October 2001; accepted 1 February 2002.

- Oró, J. Comets and the formation of biochemical compounds on the primitive Earth. *Nature* **190**, 389–390 (1961).
- Cronin, J. R. & Pizzarello, S. Amino acids in meteorites. *Adv. Space Res.* **3**, 5–18 (1983).
- Cronin, J. R. & Chang, S. in *Chemistry of Life's Origins* (NATO ASI) (eds Greenberg, J. M., Pirronello, V. & Mendoza-Gomez, C.) 209–258 (Kluwer, Dordrecht, 1993).
- Peltzer, E. T., Bada, J. L., Schlesinger, G. & Miller, S. L. The chemical conditions on the parent body of the Murchison meteorite: Some conclusions based on amino, hydroxy and dicarboxylic acids. *Adv. Space Res.* **4**, 69–74 (1984).
- Robert, F. & Epstein, S. The concentration and isotopic composition of hydrogen, carbon and nitrogen in carbonaceous meteorites. *Geochim. Cosmochim. Acta* **46**, 81–95 (1982).
- Lerner, N. R. Influence of Murchison or Allende minerals on hydrogen-deuterium exchange of amino acids. *Geochim. Cosmochim. Acta* **59**, 1623–1631 (1995).
- Lerner, N. R. Influence of Allende minerals on deuterium retention of products of the Strecker synthesis. *Geochim. Cosmochim. Acta* **61**, 4885–4893 (1997).
- Cronin, J. R. & Pizzarello, S. Enantiomeric excesses in meteoritic amino acids. *Science* **275**, 951–955 (1997).
- Sandford, S. A. The inventory of interstellar materials available for the formation of the Solar System. *Meteorit. Planet. Sci.* **31**, 449–476 (1996).
- Gibb, E. L. *et al.* An inventory of interstellar ices toward the embedded protostar W33A. *Astrophys. J.* **536**, 347–356 (2000).
- Gerakines, P. A., Schutte, W. A. & Ehrenfreund, P. Ultraviolet processing of interstellar ice analogs. I. Pure ices. *Astron. Astrophys.* **312**, 289–305 (1996).
- Tegler, S. C. *et al.* Detection of the 2165 inverse centimeter (4.619 micron) XCN band in the spectrum of L1551 IRS 5. *Astrophys. J.* **411**, 260–265 (1993).
- Bernstein, M. P., Sandford, S. A., Allamandola, L. J., Chang, S. & Scharberg, M. A. Organic compounds produced by photolysis of realistic interstellar and cometary ice analogs containing methanol. *Astrophys. J.* **454**, 327–344 (1995).
- Kerridge, J. F. Formation and processing of organics in the early Solar System. *Space Sci. Rev.* **90**, 275–288 (1999).
- Teixeira, T. C., Devlin, J. P., Buch, V. & Emerson, J. P. Discovery of solid HDO in grain mantles. *Astron. Astrophys.* **347**, L19–L22 (1999).
- Turner, B. E. Deuterated molecules in translucent and dark clouds. *Astrophys. J. Suppl. Ser.* **136**, 579–629 (2001).
- Lacy, J. H., Faraji, H., Sandford, S. A. & Allamandola, L. J. Unravelling the 10 µm 'Silicate' feature of protostars: the detection of frozen interstellar ammonia. *Astrophys. J.* **501**, L105–L109 (1998).
- Chiar, J. E. *et al.* The composition and distribution of dust along the line of sight towards the Galactic center. *Astrophys. J.* **537**, 749–762 (2000).
- Allamandola, L. J., Sandford, S. A., Tielens, A. G. G. M. & Herbst, T. M. Spectroscopy of dense clouds in the C-H stretching region: methanol and "diamonds". *Astrophys. J.* **399**, 134–146 (1992).
- Irvine, W. M. Spectroscopic evidence for interstellar ices in Comet Hyakutake. *Nature* **383**, 418–420 (1996).
- Allamandola, L. J., Sandford, S. A. & Valero, G. Photochemical and thermal evolution of interstellar/pre-cometary ice analogs. *Icarus* **76**, 225–252 (1988).
- Matthews, C. N. & Moser, R. E. Peptide synthesis from hydrogen cyanide and water. *Nature* **215**, 1230–1234 (1967).
- Ehrenfreund, P., Bernstein, M. P., Dworkin, J. P., Sandford, S. A. & Allamandola, L. J. The photostability of amino acids in space. *Astrophys. J.* **550**, L95–L99 (2001).
- Sorrell, W. H. Origin of amino acids and organic sugars in interstellar clouds. *Astrophys. J.* **555**, L129–L132 (2001).
- Rubenstein, E., Bonner, W. A., Brown, G. S. & Bailey, J. Polarized stellar light. *Science* **283**, 1415 (1999).
- Rubenstein, E., Bonner, W. A., Noyes, H. P. & Brown, G. S. Supernovae and life. *Nature* **306**, 118 (1983).
- Bonner, W. A. & Bean, B. D. Asymmetric photolysis with elliptically polarized light. *Orig. Life Evol. Biosphere* **30**, 513–517 (2000).
- Warneck, P. A microwave-powered hydrogen lamp for vacuum ultraviolet photochemical research. *Appl. Opt.* **1**, 721–726 (1962).
- Prasad, S. S. & Tarafdar, S. P. UV radiation field inside dense clouds—Its possible existence and chemical implications. *Astrophys. J.* **267**, 603–609 (1983).
- Zhao, M. & Bada, J. L. Determination of alpha-dialkylamino acids and their enantiomers in geological samples by high-performance liquid chromatography after derivatization with a chiral adduct of o-phthalaldehyde. *J. Chromatogr. A* **690**, 55–63 (1995).

Acknowledgements

This work was supported by NASA grants from the Origins of Solar Systems, Exobiology, and Astrobiology programmes, as well as the NASA Ames Director's Discretionary Fund. We thank A. Weber, D. Glavin, O. Botta, J. Chambers and K. Nelson for discussions. We also acknowledge technical support from R. Walker.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P.B. (e-mail: mbernstein@mail.arc.nasa.gov).

Amino acids from ultraviolet irradiation of interstellar ice analogues

G. M. Muñoz Caro*, U. J. Meierhenrich†‡, W. A. Schutte*, B. Barbier†, A. Arcones Segovia*, H. Rosenbauer§, W. H.-P. Thiemann‡, A. Brack† & J. M. Greenberg*

* Raymond and Beverly Sackler Laboratory for Astrophysics at Leiden Observatory, PO Box 9513, NL-2300 RA Leiden, The Netherlands

† Centre de Biophysique Moléculaire, Rue Charles Sadron, F-45160 Orléans, France

‡ Bremen University, Department of Physical Chemistry, FB 02, Leobener Straße, D-28359 Bremen, Germany

§ Max-Planck-Institut für Aeronomie, Max-Planck-Straße 2, D-37189 Katlenburg-Lindau, Germany

Amino acids are the essential molecular components of living organisms on Earth, but the proposed mechanisms for their spontaneous generation have been unable to account for their presence in Earth's early history¹. The delivery of extraterrestrial organic compounds has been proposed as an alternative to generation on Earth^{2–5}, and some amino acids have been found in several meteorites^{6–9}. Here we report the detection of amino acids in the room-temperature residue of an interstellar ice analogue that was ultraviolet-irradiated in a high vacuum at 12 K. We identified 16 amino acids; the chiral ones showed enantiomeric separation. Some of the identified amino acids are also found in meteorites. Our results demonstrate that the spontaneous generation of amino acids in the interstellar medium is possible, supporting the suggestion that prebiotic molecules could have been delivered to the early Earth by cometary dust, meteorites or interplanetary dust particles.

Dense interstellar clouds are the birthplaces of stars and planetary systems. Here, dust particles accrete ice layers with H₂O, CO, CO₂, CH₃OH and NH₃ as the main molecular components¹⁰. This ice can undergo considerable processing by stellar ultraviolet (UV) photons and cosmic rays^{11,12}. In star-forming regions it enters circumstellar disks where formation of planetary bodies as well as comets takes place. We simulated in the laboratory the processing going on in such regions.

Our experimental set-up consists of a high-vacuum chamber where a gas mixture is deposited on a cold 'finger' (an aluminium block at 12 K) and irradiated by UV (ref. 13). An ice mixture containing H₂O:CH₃OH:NH₃:CO:CO₂ = 2:1:1:1:1 (molar composition) was selected as representative of the interstellar medium. This composition resembles the ice found close to protostellar sources^{10,14,15}. After warming the system to room temperature a small amount of material remained. The dominant components of

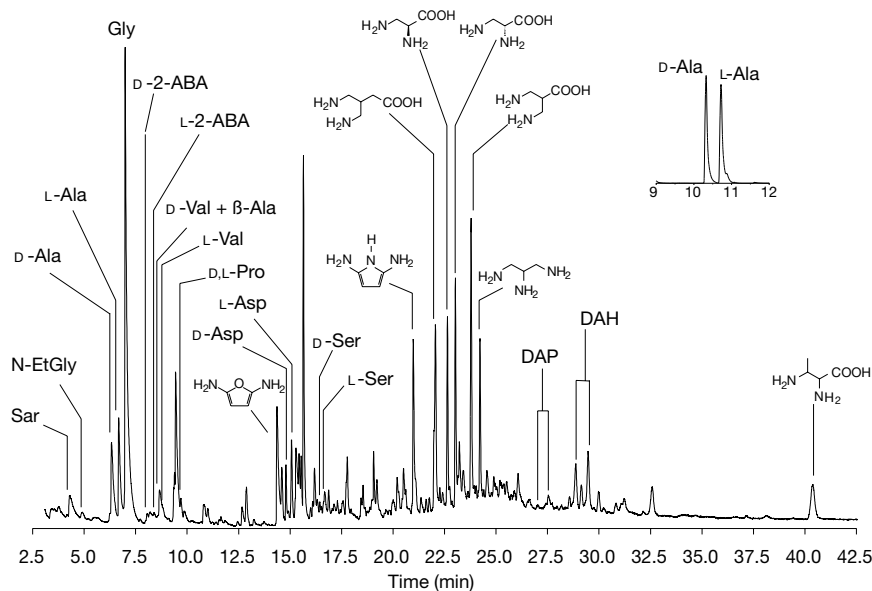


Figure 1 Gas chromatogram showing the amino acids and other compounds generated under simulated interstellar pre-cometary conditions. Data were obtained from analysis of the room temperature residue of photoprocessed interstellar medium ice analogue taken after 6 M HCl hydrolysis and derivatization (ECEE derivatives, Varian-Chrompack Chirasil-L-Val capillary column 12 m $\times$ 0.25 mm inner diameter, layer thickness 0.12 μ m; splitless injection, 1.5 ml min⁻¹ constant flow of He carrier gas; oven temperature

programmed for 3 min at 70 °C, 5 °C min⁻¹, and 17.5 min at 180 °C; detection of total ion current with GC-MSD system Agilent 6890/5973). The inset shows the determination of alanine enantiomers in the above sample (Chirasil-L-Val 25 m, single ion monitoring for Ala-ECEE base peak at 116 a.m.u.). DAP, diaminopentanoic acid; DAH, diaminohexanoic acid; a.m.u., atomic mass units.

this residue were saturated organic compounds containing carboxylic groups and hexamethylenetetramine^{16–18}. Previous work claimed the detection of trace amounts of a limited number of amino acids in photolysed interstellar ice analogues^{17,19}. Except for glycine, however, these analytical results were not confirmed by isotopic labelling, a necessary verification to exclude contamination with trace organic components.

We analysed the residue by gas chromatography-mass spectrometry (GC-MS) using a chiral capillary column. Using new, more sensitive detection methods, we report here the identification of 16 amino acids in the residue, 6 of which are protein constituents. Figure 1 shows the chromatogram corresponding to the residue obtained from the photoprocessed interstellar ice analogue mixture. Most of the observed peaks were due to amino acids, but three exceptions were identified as 2,5-diaminopyrrole, 2,5-diaminofuran, and 1,2,3-triaminopropane. The calculated signal-to-noise ratio was $S/N \geq 10$, even for the least abundant amino acids identified. A similar chromatogram was obtained for the isotopically labelled sample (molar concentration $\text{H}_2\text{O}^{13}\text{CH}_3\text{OH}:\text{NH}_3^{13}\text{CO}^{13}\text{CO}_2 = 2:1:1:1$), with the exceptions of valine and proline, which were below the detection limit. Proline was found in the ^{13}C -labelled residue of a test sample, which did not contain CH_3OH in the starting ice composition, and therefore could not be a contaminant. The abundance was obtained by integrating the peak area of a given amino acid calibrated with the corresponding standard. The quantum yield (Φ) was obtained by dividing the abundance by the number of photons. For the total amount of residue, $\Phi(\text{residue}) \approx 8.0 \times 10^{-3}$ (ref. 20). Glycine, the simplest amino acid, with $\Phi(\text{Gly}) \approx 3.6 \times 10^{-5}$, was the most abundant. The total amino-acid quantum yield was $\Phi \approx 1.0 \times 10^{-4}$. Table 1 lists the detected molecules, and their corresponding abundances relative to glycine. The mass fragments of the elutants in the ^{12}C and ^{13}C samples are given along with their retention times (R_t). The main mass fragments are shown in boldface in Table 1. For amino acids bearing only one amino group, species with higher molecular mass were less abundant. We were surprised to find that species with two amino

groups showed relatively high abundances. Aspartic acid was the only dicarboxylic compound found in the residues. Decarboxylation is known as the main photodamage of UV-exposed amino acids and probably competes with photoproduction²¹. Decarboxylation should give rise to the formation of radicals that could react with NH_3 to favour the formation of diamino-compounds.

In the case of reported analyses of meteoritic material, the absolute abundances of the detected amino acids are enhanced by acid hydrolysis⁶. Similarly in our case, amino acids in the residue are detected at considerable amounts only after acid hydrolysis, which probably indicates that these products are originally formed as peptidic molecules, and released as free amino acids after hydrolysis at room temperature. It has been suggested that hydrolysis of hexamethylenetetramine, an abundant product of photolysis for the selected ice mixture, leads to formation of amino acids¹⁸. In order to test this possibility, an external standard of hexamethylenetetramine was analysed using the same protocol as for the residues and no amino acids were detected. No evidence for contamination was found after comparison with the samples produced from ^{13}C -labelled reactants. We used the mass spectrometric data on amino-acid ECEE (N-ethoxycarbonyl ethyl ester) derivatives reported in the literature²² for identification.

The selected starting ice composition here reported is only indicative, because the composition of the ice can vary significantly for different sources. The NH_3 ice abundance used was enhanced with respect to that derived from astronomical observations to obtain a larger amount of residue, but it is also based on the fact that NH_3 photo-dissociates very readily, leading to other species, so that the original budget of NH_3 was probably larger than what is observed. Additional irradiation experiments involving a ten times higher H_2O ice concentration or deposition at 80 K, monitored by infrared spectroscopy, showed that the residue production was equally efficient, although the relative abundances of the main components, hexamethylenetetramine and carboxylic acids, varied somewhat²³.

All the chiral amino acids detected showed enantiomeric separa-

Table 1 Peak identification for Fig. 1

Amino acid	Quantum yield of ISM sample, $\Phi \times 100/\Phi$ (Gly)	MS-fragmentation		R_t of analyte (min)	Biological occurrence
		^{12}C -sample (a.m.u.)	^{13}C -sample (a.m.u.)		
Glycine	100	175, 130, 102	177, 132, 103	6.99	Yes
α -D-alanine	19.3	189, 144, 116 , 88	147, 118 , 90	6.33	No
α -L-alanine	20.0	189, 144, 116 , 88	147, 118 , 90	6.68	Yes
β -alanine	4.29	189, 160, 144, 116, 115 , 102, 98	192, 147, 119, 117 , 103, 101	8.66	No
Sarcosine (N-methylglycine)	5.71	189, 144, 116 , 88	192, 147, 118 , 90	4.29	No
D-2-aminobutyric acid	0.46	130	133	8.05	No
L-2-aminobutyric acid	0.48	130	133	8.37	No
N-ethylglycine	1.91	130 , 84, 58	133 , 207	4.90	No
D-valine	0.61	144	$\leq$ d.l.	8.66	No
L-valine	0.61	144	$\leq$ d.l.	8.80	Yes
D,L-proline	0.06	142	$\leq$ d.l.	9.59	No, Yes
D-serine	3.29	175 (McLafferty), 160, 132 , 114	116, 134	16.45	No
L-serine	3.86	204, 187, 175 (McLafferty), 160, 132 , 114	116, 134	16.67	Yes
D-aspartic acid	1.14	188 , 142	191 , 145	14.90	No
L-aspartic acid	1.07	188 , 142	191	15.05	Yes
D-2,3-diaminopropanoic acid	20.0	231, 203, 188, 175 , 157, 129, 102	234, 205, 190, 177 , 159, 131, 103	22.64	No
L-2,3-diaminopropanoic acid	20.0	231, 203, 188, 175 , 157, 129, 102	234, 205 190, 177 , 159, 131, 103	23.03	No
2,3-diaminobutyric acid	14.1	203 , 157, 115, 85	205	40.38	No
3,3'-diaminoisobutyric acid	32.9	290, 245, 238, 216, 188, 167, 156, 142 , 128, 115, 102	249, 219, 191, 145 , 132, 103	23.79	No
4,4'-diaminoneopentanoic acid	22.4	304, 259, 230, 202, 188, 182, 166, 156, 142, 116	264, 206, 191, 187, 171, 160, 145, 118	22.05	No
A-diaminopentanoic acid*	1.72	142 (c.i.i.), 129, 70	$\leq$ d.l.	27.05	No
B-diaminopentanoic acid*	1.92	142 (c.i.i.)	$\leq$ d.l.	27.54	No
A-diaminohexanoic acid†	33.1	272, 156 (c.i.i.), 128, 102, 84	272(cyclic ion), 160	28.86	No
B-diaminohexanoic acid†	35.4	272, 156 (c.i.i.), 128, 102, 84	272(cyclic ion), 160	29.48	No
2,5-diaminofurane	23.1	242, 169, 153, 125, 97 , 81, 69, 53	246, 173, 157, 129, 101 , 85, 72, 56	14.36	
2,5-diaminopyrrole	20.0	241, 213, 212, 196, 168 , 140, 124, 109, 96 , 81, 69, 54, 53	245, 217, 216, 200, 172 , 144, 128, 113, 100, 85, 72, 57	20.98	
1,2,3-triaminopropane	16.3	232, 187, 159, 144, 130, 115, 102	189, 162, 147, 132, 117, 103	24.22	
Unidentified	36.5	233, 188, 160, 129, 114, 102	237, 192, 163, 131, 117, 103	15.64	

For glycine, $\Phi(\text{Gly}) \approx 3.6 \times 10^{-5}$. Numbers designated in boldface are the main mass fragments. A designates the first eluting enantiomer, and B the second eluting enantiomer. c.i.i., cyclic immonium ion. R_t , retention time. ISM, interstellar medium. MS, mass spectrometry. d.l., detection limit; McLafferty indicates that a McLafferty rearrangement has occurred.

*Exact molecular structure not yet known. Ornithine or its constitutional isomer.

†Exact molecular structure not yet known. Lysine or its constitutional isomer.

tion, with the exception of proline, which could not be separated by the selected protocol²⁴. As expected for photosynthesis using non-polarized light, the amino acids produced appeared to be racemic, with enantiomeric excesses fluctuating within experimental errors.

Using the above efficient derivatization and sensitive GC-MS technique we were able to identify a total of 16 different amino acids in the residue. Hence, we show that photochemical processing of interstellar ice analogues leads to a large variety of products, which can only be identified by very specific detection techniques. We recently found evidence for the presence of a number of new organic components in the residues²³.

In addition to amino acids, 1,2,3-triaminopropane, 2,5-diaminofurane and 2,5-diaminopyrrole were identified as shown in Fig. 1. This is the first detection of pyrroles and furanes in photolysed interstellar ice analogues. Such cyclic molecules were tentatively identified in the mass spectra of comet Halley²⁵. All the products reported in Table 1 were also found in the residue of a sample containing only CO as the carbon carrier (starting ice molar composition $\text{H}_2\text{O}:\text{NH}_3:\text{CO} = 2:1:1$). We also obtained results from experiments involving different ice mixtures²³.

Glycine has so far not been detected in the interstellar medium. Some of the amino acids reported here were also found in carbonaceous chondrites^{6–9}. However, no amino acids with more than one amino group were detected in these objects. Our work suggests that these species might also be present in meteorites.

How life originated is one of the earliest and most intriguing questions for humanity. Early experiments on the processing of a gas mixture simulating the primitive Earth conditions assumed a reducing atmosphere with CH_4 as the carbon-containing molecule²⁶. Several amino acids were formed under these conditions as products of spark discharge, photoprocessing or heat. It is now believed, however, that the Earth's early atmosphere was rather non-reducing, with CO_2 as the main carbon carrier. Processing of these

alternative gas mixtures under experimental conditions leads to the formation of, at most, traces of amino acids¹.

The results reported in this paper support the concept of extraterrestrial delivery of prebiotic molecules triggering the appearance of life on Earth^{2–5}. We show that, at least to some extent, amino acids and other prebiotic molecules such as pyrroles and furanes could be of extraterrestrial origin. Earlier analysis of the residues of photochemically processed interstellar ice analogues revealed other components of prebiotic interest such as quinones and aromatic alcohols²⁷. It is probable that comets, to a large extent, consist of pristine interstellar materials^{28,29}. Such species could, therefore, have been delivered to the early Earth by comet dust, as well as asteroids and interplanetary dust particles during the epoch of heavy bombardment³⁰. Similar experiments using circularly polarized ultraviolet light could result in the formation of non-racemic amino acids, such as those found in the proteins of living organisms, which are exclusively of L-configuration.

Several parameters still need to be better constrained (such as UV exposure, particle fluxes, and number of collisions) before a reliable estimation on the extraterrestrial delivery of amino acids to the early Earth can be made. To this end, *in situ* analysis of cometary material will be performed in the near future by space probes such as Rosetta and Stardust. □

Methods

Simulation of the interstellar medium

The high-vacuum system for UV-irradiation of the ice was powered by a turbo pump (Pfeiffer Balzers TSH 280H) backed up by a diaphragm pump (Vacuubrand MD4T). The system pressure at room temperature was $P_s \approx 10^{-7}$ mbar. A temperature of 12 K, typical of dust in dense clouds, was achieved by means of a closed-cycle helium cryostat (Air Products Displex DE-202). CO_2 was kept in a separate bulb in order to prevent it from reacting with the ammonia, NH_3 . Simultaneous deposition of CO_2 and the other components was achieved via two separate deposition tubes, one for each bulb. The gas molecules were condensed on an aluminium block mounted on the tip of the cryostat.

During the deposition, the accreting ice layer was irradiated by UV from a microwave-stimulated hydrogen flow discharge lamp (output, $F \approx 1.5 \times 10^{15}$ photons s^{-1} (ref. 31); photon energy, $E_{\text{photon}} = 7.3\text{--}10.5$ eV, with main emission at Lyman- α (10.2 eV)). The lamp spectrum roughly resembles the UV emission from early-type stars, which dominates the interstellar radiation field between 0.09 and 0.25 μm (ref. 20). The UV-irradiation/deposition typically lasted 24 hours. The total gas flow was $F \approx 1 \times 10^{16}$ molecules s^{-1} , implying an average UV dose of 0.15 photons per molecule. Subsequently, the system was warmed at 1 K min^{-1} to $T = 40 \text{ K}$ using a temperature controller (Scientific Instruments 9600-1), and then at about 4 K min^{-1} up to room temperature. The aluminium block with the residue was removed from the system and stored inside a capsule in a nitrogen atmosphere.

Analytical procedure

A 100- μl water extract of the above produced residue was hydrolysed in 6 M HCl at 110°C for 24 h under argon atmosphere. This is a standard procedure for hydrolysis of peptides and has been used for the detection of amino acids in meteorites⁷. After evaporation of 6 M HCl, the residues were dissolved in 0.1 M HCl and derivatized using the procedure of ref. 24, leading to amino acid ECEE derivatives. The identities of the amino acid peaks were verified by comparing the retention times and the mass spectra with external standards, purchased from Fluka. By using the single ion monitoring mode, it was possible to select a given ion mass, characteristic of the target amino acid in order to increase both chromatographic resolution and sensitivity. The determination of the signal-to-noise ratio for each amino acid was performed in the extract ion mode.

Received 16 November 2001; accepted 17 January 2002.

- Bar-Nun, A. & Chang, S. Photochemical reactions of water and carbon monoxide in Earth's primitive atmosphere. *J. Geophys. Res.* **88**, 6662–6672 (1983).
- Oró, J. Comets and the formation of biochemical compounds on the primitive Earth. *Nature* **190**, 389–390 (1961).
- Greenberg, J. M. in *The Galaxy and the Solar System* 103–115 (Univ. Arizona Press, Tucson, 1986).
- Brack, A. Life in the Solar System. *Adv. Space. Res.* **24**, 417–433 (1999).
- Meierhenrich, U., Thiemann, W.H.-P. & Rosenbauer, H. Molecular parity violation via comets? *Chirality* **11**, 575–582 (1999).
- Engel, M. H. & Macko, S. A. The stereochemistry of amino acids in the Murchison meteorite. *Precamb. Res.* **106**, 35–45 (2001).
- Kvenvolden, K. A. *et al.* Evidence for extraterrestrial amino acids and hydrocarbons in the Murchison meteorite. *Nature* **228**, 923–926 (1970).
- Cronin, J. R. & Pizzarello, S. Enantiomeric excesses in meteoritic amino acids. *Science* **275**, 951–955 (1997).
- Ehrenfreund, P., Glavin, D. P., Botta, O., Cooper, G. & Bada, J. L. Extraterrestrial amino acids in Orgueil and Ivuna: Tracing the parent body of CI type carbonaceous chondrites. *Proc. Natl Acad. Sci. USA* **98**, 2138–2141 (2001).
- Gibb, E. L., Whittet, D. C. & Chiar, J. E. Searching for ammonia in grain mantles toward massive young stellar objects. *Astrophys. J.* **558**, 702–716 (2001).
- Muñoz Caro, G. M., Ruiterkamp, R., Schutte, W. A., Greenberg, J. M. & Mennella, V. UV photodestruction of CH bonds and the evolution of the 3.4 μm feature carrier. I. The case of aliphatic and aromatic molecular species. *Astron. Astrophys.* **367**, 347–354 (2001).
- Mennella, V. *et al.* UV photodestruction of CH bonds and the evolution of the 3.4 μm feature carrier. II. The case of hydrogenated carbon grains. *Astron. Astrophys.* **367**, 355–361 (2001).
- Gerakines, P. A., Schutte, W. A., Greenberg, J. M. & van Dishoeck, E. F. The infrared band strengths of H_2O , CO and CO_2 in laboratory simulations of astrophysical ice mixtures. *Astron. Astrophys.* **296**, 810–818 (1995).
- Gerakines, P. A. *et al.* Observations of solid carbon dioxide in molecular clouds with the infrared space observatory. *Astrophys. J.* **522**, 357–377 (1999).
- Ehrenfreund, P. *et al.* Laboratory studies of thermally processed $\text{H}_2\text{O}:\text{CH}_3\text{OH}:\text{CO}_2$ ice mixtures and their astrophysical implications. *Astron. Astrophys.* **350**, 240–253 (1999).
- Greenberg, J. M., Yench, A. J., Corbell, J. W. & Frisch, H. L. in *Extrait de Mémoires de la Société Royale des Sciences de Liège* 425 (Société Royale des Sciences de Liège, Liège, 1972).
- Briggs, R. *et al.* Comet Halley as an aggregate of interstellar dust and further evidence for the photochemical formation of organics in the interstellar medium. *Orig. Life Evol. Biosphere* **22**, 287–307 (1992).
- Bernstein, M. P., Sandford, S. A., Allamandola, L. J., Chang, S. & Scharberg, M. A. Organic compounds produced by photolysis of realistic interstellar and cometary ice analogs containing methanol. *Astrophys. J.* **454**, 327–344 (1995).
- Kobayashi, K. *et al.* Synthesis of amino acids in Earth orbit: Proposal. *Adv. Space Res.* **23**, 401–404 (1999).
- Jenniskens, P. *et al.* Carbon dust formation on interstellar grains. *Astron. Astrophys.* **273**, 583–600 (1993).
- Johns, R. B. & Seuret, M. G. Photochemistry of biological molecules. III. Mechanism of photodamage of alanine peptides in the solid state. *Photochem. Photobiol.* **12**, 405–417 (1970).
- Huang, Z.-H., Wang, J., Gage, D. A., Watson, J. T. & Sweeley, C. C. Characterization of N-ethoxycarbonyl ethyl esters of amino acids by mass spectrometry. *J. Chromatogr.* **635**, 271–281 (1993).
- Muñoz Caro, G. M., Meierhenrich, U. J., Schutte, W. A. & Greenberg, J. M. UV photoprocessing of interstellar ice analogs. Formation of HMT derivatives. *Astron. Astrophys.* (submitted).
- Abe, I., Fujimoto, N., Nishiyama, T., Terada, K. & Nakahara, T. Rapid analysis of amino acid enantiomers by chiral-phase capillary gas chromatography. *J. Chromatogr. A* **722**, 221–227 (1996).
- Kissel, J. & Krueger, F. R. The organic component in dust from comet Halley as measured by the PUMA mass spectrometer on board Vega 1. *Nature* **326**, 755–760 (1987).
- Miller, S. L. A production of amino acids under possible primitive Earth conditions. *Science* **117**, 528–529 (1953).
- Bernstein, M. P. *et al.* UV irradiation of polycyclic aromatic hydrocarbons in ices: production of alcohols, quinones, and ethers. *Science* **283**, 1135–1138 (1999).
- Greenberg, J. M. in *Comets* (ed. Wilkening, L. L.) 131–163 (Univ. Arizona Press, Tucson, 1982).

- Irvine, W. M., Schloerb, F. P., Crovisier, J., Fegley, B. Jr & Mumma, M. J. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 1159–1200 (Univ. Arizona Press, Tucson, 2000).
- Chyba, C. & Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**, 125–131 (1992).
- Weber, P. & Greenberg, J. M. Can spores survive in interstellar space? *Nature* **316**, 403–407 (1985).

Acknowledgements

We thank A. MacDermott, F. Goesmann and R. Roll for discussions and K. Getliffe for suggestions. Our collaboration was originally supported by the Max-Planck-Institut für Aeronomie at Katlenburg-Lindau with grants for G.M.M.C. and U.J.M., as preparatory work for the Rosetta mission. The work was performed at the Sackler Laboratory for Astrophysics at Leiden Observatory and the Centre Biophysique Moléculaire at Orléans, where it was funded by the European Union and by French sponsors: CNES, FNAD, Région Centre, and Département du Cher. In Orléans U.J.M. was supported by a Le Studium grant. U.J.M. is grateful for a habilitation grant by the Deutsche Forschungsgemeinschaft, and G.M.M.C. for an AIO allowance by Leiden University. This paper is dedicated to the memory of J.M.G., who died on 29 November 2001.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.J.M. (e-mail: mhenrich@uni-bremen.de).

Exchange-biased quantum tunnelling in a supramolecular dimer of single-molecule magnets

Wolfgang Wernsdorfer*, Núria Aliaga-Alcalde†, David N. Hendrickson‡ & George Christou†

* Laboratoire Louis Néel-CNRS, BP166, 25 Avenue des Martyrs, 38042 Grenoble, Cedex 9, France

† Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200, USA

‡ Department of Chemistry, University of California at San Diego, La Jolla, California 92093-0358, USA

Various present and future specialized applications of magnets require monodisperse, small magnetic particles, and the discovery of molecules that can function as nanoscale magnets was an important development in this regard^{1–3}. These molecules act as single-domain magnetic particles that, below their blocking temperature, exhibit magnetization hysteresis, a classical property of macroscopic magnets. Such ‘single-molecule magnets’ (SMMs)⁴ straddle the interface between classical and quantum mechanical behaviour because they also display quantum tunnelling of magnetization^{5,6} and quantum phase interference⁷. Quantum tunnelling of magnetization can be advantageous for some potential applications of SMMs, for example, in providing the quantum superposition of states required for quantum computing⁸. However, it is a disadvantage in other applications, such as information storage, where it would lead to information loss. Thus it is important to both understand and control the quantum properties of SMMs. Here we report a supramolecular SMM dimer in which antiferromagnetic coupling between the two components results in quantum behaviour different from that of the individual SMMs. Our experimental observations and theoretical analysis suggest a means of tuning the quantum tunnelling of magnetization in SMMs. This system may also prove useful for studying quantum tunnelling of relevance to mesoscopic antiferromagnets.

The compound $[\text{Mn}_4\text{O}_3\text{Cl}_4(\text{O}_2\text{Cet})_3(\text{py})_3]$ (hereafter Mn4) con-

tains three Mn^{3+} ions and one Mn^{4+} ion, and exchange coupling between them leads to Mn4 having a ground state spin of $S = 9/2$. Mn4 crystallizes in the hexagonal space group $R\bar{3}$ with pairs of Mn4 molecules lying 'head to head' on a crystallographic S_6 symmetry axis (Fig. 1)⁹. Thus, each Mn4 has C_3 symmetry and the $[\text{Mn4}]_2$ dimer has S_6 symmetry. This supramolecular arrangement is held together by six $\text{C}-\text{H}\cdots\text{Cl}$ hydrogen bonds (dashed lines in Fig. 1) between the pyridine rings on one $[\text{Mn4}]$ and Cl ions on the other. $\text{C}-\text{H}\cdots\text{X}$ (where X is an electronegative atom) hydrogen bonds^{10–12} are weaker than their $\text{O}-\text{H}\cdots\text{X}$ and $\text{N}-\text{H}\cdots\text{X}$ analogues but are nevertheless of fundamental importance in the packing of organic and inorganic molecules in the solid state, the folding of biomolecules, and a wide variety of molecular recognition events, crystal engineering, and similar types of interactions^{13–15}, particularly when X is O . The $\text{C}-\text{H}\cdots\text{Cl}$ hydrogen bonds in $[\text{Mn4}]_2$ have $\text{C}\cdots\text{Cl}$ distances and $\text{C}-\text{H}\cdots\text{Cl}$ angles of $3.71 \text{ \AA}$ and 161.57° , respectively, typical of this type of hydrogen bond^{11,16,17}. The $[\text{Mn4}]_2$ dimer also brings the central bridging Cl ions of each $[\text{Mn4}]$ close together ($3.858 \text{ \AA}$, dotted line in Fig. 1), almost at their van der Waals separation ($\sim 3.6 \text{ \AA}$). Each $[\text{Mn4}]_2$ dimer is well separated from neighbouring dimers.

The supramolecular linkage within $[\text{Mn4}]_2$ introduces exchange interactions between the Mn4 molecules via both the six $\text{C}-\text{H}\cdots\text{Cl}$

pathways and the $\text{Cl}\cdots\text{Cl}$ approach. Most of the spin density at each Mn^{3+} ion is in d_π -orbitals, and some of this is known^{9,18} to be delocalized into the pyridine π -system. Overlap with Cl π -orbitals of the neighbour provides an antiferromagnetic interaction. Similarly, the approach of the central Cl ions provides another pathway for antiferromagnetic exchange; propagation of the latter through metal-bound Cl ions is extremely well documented¹⁹. All seven interactions in $[\text{Mn4}]_2$ are expected to be weak, but combined they will lead to noticeable antiferromagnetic coupling between the Mn4 units.

The antiferromagnetic coupling between the Mn4 units in $[\text{Mn4}]_2$ makes this dimer an excellent candidate for studying quantum tunnelling in a system of truly identical, antiferromagnetically coupled particles with $S = 0$ ground states. Tunnelling studies were performed by magnetization measurements on single crystals using an array of micro-SQUIDS²⁰. Figure 2 shows typical hysteresis loops in magnetization versus magnetic field scans with the field applied approximately along the easy axis of magnetization of $[\text{Mn4}]_2$, that is, approximately parallel to the S_6 axis. These loops display step-like features separated by plateaus. The step heights are temperature-dependent above 0.3 K ; below this, the loops become temperature-independent. As discussed below, the steps are due to resonant quantum tunnelling of the magnetization (QTM) between the energy states of the $[\text{Mn4}]_2$ dimer. QTM has been previously observed for several SMMs^{5–7,20–25}, but what makes $[\text{Mn4}]_2$ unusual is that the QTM is now the collective behaviour of the complete $S = 0$ dimer of exchange-coupled $S = 9/2$ Mn4 quantum systems. This coupling is manifested as an exchange bias of all tunnelling transitions, and the hysteresis loops consequently display unique features, such as the absence for the first time in a SMM of a QTM step at zero field.

Before interpreting the hysteresis loops further, we present a

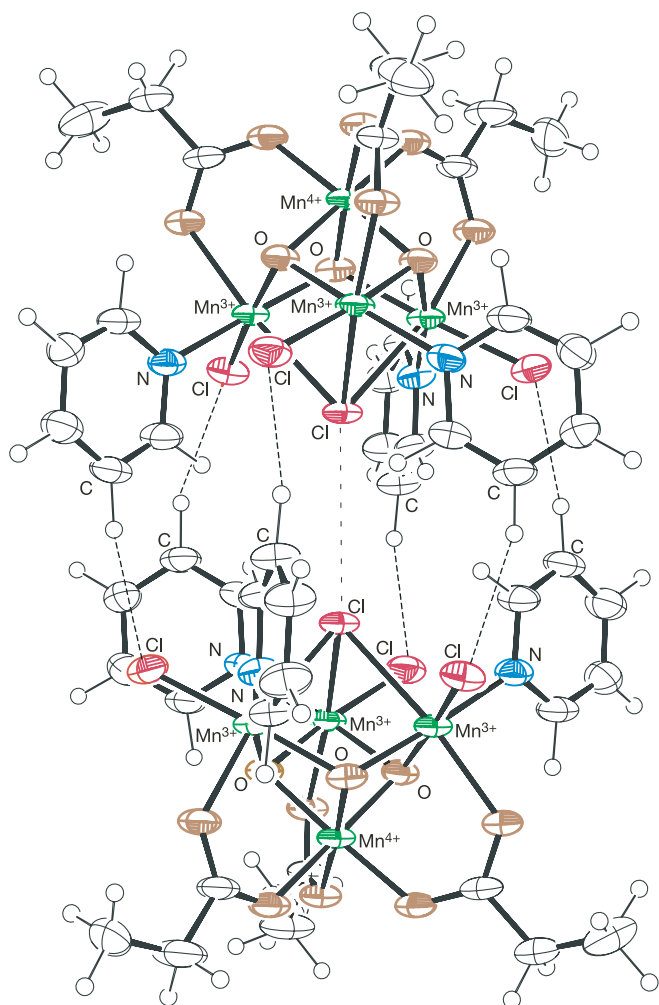


Figure 1 The structure of the $[\text{Mn}_4\text{O}_3\text{Cl}_4(\text{O}_2\text{CEt})_3(\text{py})_3]_2$ dimer, denoted $[\text{Mn4}]_2$. The small circles are hydrogen atoms. The dashed lines are $\text{C}-\text{H}\cdots\text{Cl}$ hydrogen bonds and the dotted line is the close $\text{Cl}\cdots\text{Cl}$ approach. Brown, oxygen; green, manganese; red, chlorine; blue, nitrogen; py, pyridine.

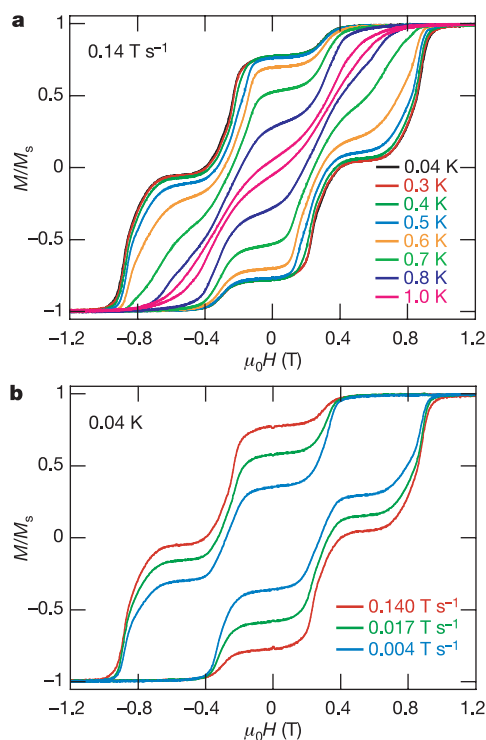


Figure 2 Magnetization (M) of $[\text{Mn4}]_2$ (plotted as fraction of maximum value M_s) versus applied magnetic field ($\mu_0 H$). **a**, **b**, The resulting hysteresis loops are shown at different temperatures (**a**) and different field sweep rates (**b**). We note that the loops become temperature-independent below about 0.3 K but are still sweep-rate-dependent owing to resonant quantum tunnelling between discrete energy states of the $[\text{Mn4}]_2$ dimer.

simplified spin hamiltonian describing the $[\text{Mn4}]_2$ dimer. Each Mn4 can be modelled as a 'giant spin' of $S = 9/2$ with Ising-like anisotropy. The corresponding hamiltonian (H_i) is given by equation (1)

$$H_i = D\hat{S}_{zi}^2 + H_i^{\text{trans}} + g\mu_B\mu_0\hat{S}_{zi}H_z \quad (1)$$

where $i = 1$ or 2 (referring to the two Mn4 SMMs of the dimer), D is the axial anisotropy constant, μ_B is the Bohr magneton, $\hat{S}_z$ is the easy-axis spin operator, g is the electronic g -factor, μ_0 is the vacuum permeability, and H_z is the applied longitudinal field. The exact form of the transverse anisotropy H_i^{trans} is not important in this discussion. The last term in equation (1) is the Zeeman energy associated with an applied field. The Mn4 units within $[\text{Mn4}]_2$ are coupled by a weak superexchange, J , via both the six C–H...Cl pathways and the Cl...Cl approach. Thus, the hamiltonian (H) for $[\text{Mn4}]_2$ is

$$H = H_1 + H_2 + J\hat{S}_1\hat{S}_2 \quad (2)$$

where $S_1 = S_2 = 9/2$. Tunnelling among the $(2S_1 + 1)(2S_2 + 1) = 100$ energy states is allowed by the small transverse anisotropy H_i^{trans} and the transverse coupling terms containing $\hat{S}_{xi}$ and $\hat{S}_{yi}$ operators. The energy states of $[\text{Mn4}]_2$ can easily be calculated by exact diagonalization; however, for simplicity, all transverse anisotropy terms were neglected. Each state of $[\text{Mn4}]_2$ can then be labelled by two quantum numbers (M_1, M_2) for the two Mn4 SMMs, with $M_1 = 9/2, 7/2, \dots, -9/2$ and

$M_2 = 9/2, 7/2, \dots, -9/2$. The energy states are plotted in Fig. 3a as a function of applied field.

At very low temperature, most of the excited spin states are not populated and can be neglected. Thus, Fig. 3b shows only the low-lying states involved in the magnetization reversal at very low temperature when sweeping the field from a high negative field to a positive one. At high negative field, the initial state is $(M_1, M_2) = (-9/2, -9/2)$, that is, both molecules are in the negative ground state. As the magnetic field is swept, the first (avoided) level crossing is at -0.33 T. There is a non-zero probability, which depends upon the sweep rate, for the magnetization to tunnel from $(-9/2, -9/2)$ to $(-9/2, 9/2)$. (For convenience, we do not list here both degenerate (M, M') and (M', M) states.) The smaller the sweep rate the larger is the tunnelling probability, that is, the larger is the resulting step in Fig. 2b. At field values away from the avoided level crossing, the dimer states are frozen by the significant magnetic anisotropy barrier.

At the next level crossing at 0 T, there is the possibility of tunnelling from $(-9/2, -9/2)$ to $(9/2, 9/2)$, but this requires both Mn4 molecules of the dimer to tunnel simultaneously. The corresponding tunnelling probability is very small and can be neglected. Between 0.15 and 0.75 T, there are several level crossings from $(-9/2, -9/2)$ to excited states (shown only in Fig. 3a) that require simultaneous tunnelling in both Mn4 units and will not be discussed further. At the next level crossing at 0.2 T, $[\text{Mn4}]_2$ can undergo tunnelling from $(-9/2, -9/2)$ to $(-9/2, 7/2)$, corresponding to one of the Mn4 molecules tunnelling from the ground state to an excited state, followed by rapid relaxation from $(-9/2, 7/2)$ to $(-9/2, 9/2)$ (curly arrow in Fig. 3b). At the next level crossing at 0.33 T, the situation is analogous to that at -0.33 T, and the dimer can undergo tunnelling from $(-9/2, 9/2)$ to $(9/2, 9/2)$. The level crossing at 0.75 T allows tunneling from $(-9/2, -9/2)$ to $(-9/2, 5/2)$, followed by relaxation from $(-9/2, 5/2)$ to $(-9/2, 7/2)$ and $(-9/2, 9/2)$. Finally, at 0.87 T tunnelling can occur from $(-9/2, 9/2)$ to $(7/2, 9/2)$, followed by relaxation from $(7/2, 9/2)$ to $(9/2, 9/2)$.

The values of D and J used in Fig. 3 were calculated from the field positions of the steps in the hysteresis loops, ignoring transverse (and fourth order) terms; the positions were determined from the first derivative (Fig. 4). We note that closely spaced peaks are not well resolved, owing to broadening by dipolar and transverse fields, and possibly other effects; this and the variation in broadnesses are under further study. The calculated values are $D = -0.72$ K and $J = +0.1$ K. The former is very close to those determined experimentally for several (isolated) Mn4 SMMs^{23,26}, and the latter is

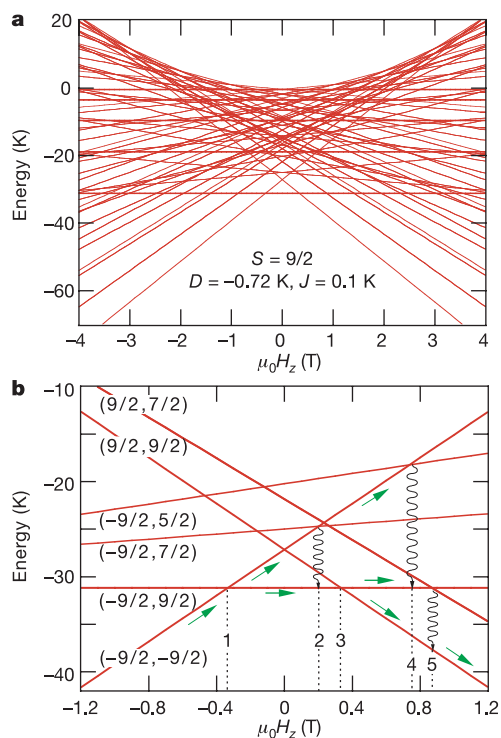


Figure 3 The spin state energies of $[\text{Mn4}]_2$ as a function of applied magnetic field. **a**, Energy versus magnetic field plot for the 100 states of the exchange-coupled dimer of two spin $S = 9/2$ Mn4 units. Transverse terms in equations (1) and (2) are neglected for the sake of simplicity. **b**, Enlargement of **a**, showing only levels populated at very low temperature when the field is swept from -1.2 T to $+1.2$ T, as indicated by green arrows. Dotted lines, labelled 1 to 5, indicate the strongest tunnel resonances: 1, $(-9/2, -9/2)$ to $(-9/2, 9/2)$; 2, $(-9/2, -9/2)$ to $(-9/2, 7/2)$, followed by relaxation to $(-9/2, 9/2)$; 3, $(-9/2, 9/2)$ to $(9/2, 9/2)$; 4, $(-9/2, -9/2)$ to $(-9/2, 5/2)$, followed by relaxation to $(-9/2, 9/2)$; 5, $(-9/2, 9/2)$ to $(7/2, 9/2)$, followed by relaxation to $(9/2, 9/2)$. For clarity, degenerate states such as (M, M') and (M', M) are not listed. S , spin value; D , axial anisotropy constant; J , superexchange parameter.

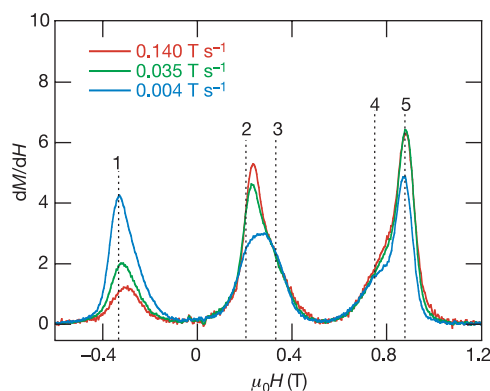


Figure 4 Derivative of the hysteresis loop at 0.04 K (Fig. 2b) and at different field sweep rates. The dominating tunnel transitions are indicated by dashed lines and numbers 1 to 5, which are indicated in Fig. 3b and explained in the text. The dashed lines are at positions calculated from Fig. 3b (which ignores transverse terms in equations (1) and (2), as well as fourth-order terms). M , magnetization; H , field.

antiferromagnetic (positive value) and very weak, as expected for an exchange interaction via the six C–H···Cl and the Cl···Cl pathways. A value of $J \approx 0.1$ K was also obtained from d.c. susceptibility measurements in the 1 to 8 K range.

The above results demonstrate that even weak exchange interactions can have a large influence of the quantum properties of SMMs. From one viewpoint, each half of the $[\text{Mn4}]_2$ dimer acts as a field bias on its neighbour, shifting the tunnel resonances to new positions relative to isolated Mn4 molecules. In particular, these single Mn4 molecules show a strong QTM step at zero field in the hysteresis loop²³, a feature absent for $[\text{Mn4}]_2$ (Fig. 2) because this would be a double quantum transition with a very low probability of occurrence. The absence of tunnelling at zero field is important if SMMs are to be used for information storage, and the option is still retained to switch on tunnelling, if and when required, by application of a field. Thus, future studies will investigate this double transition at zero field in $[\text{Mn4}]_2$, and the corresponding multiple quantum transition in higher supramolecular aggregates of SMMs, and determine their exact probability of occurrence and to what extent this can be controlled by the degree of aggregation, variation of exchange coupling strength and similar modifications. From another viewpoint, $[\text{Mn4}]_2$ represents an example of quantum tunnelling within a monodisperse antiferromagnetically coupled particle with no uncompensated spin (that is, $S = 0$) in the ground state. It may also prove possible to study its $(9/2, -9/2)$ to $(-9/2, 9/2)$ tunnelling transition. This would be analogous to the transition in antiferromagnets, in which tunnelling is predicted²⁷ to be more pronounced than in ferromagnets. Such a study for ferritin suffered from the practical impossibility of having all molecules in the sample be the same size, and possess completely compensated spins^{28–30}. Finally, the absence of a level crossing at zero field also makes $[\text{Mn4}]_2$ a very interesting candidate as a qubit for quantum computing⁸, because its ground state is the entangled combination of the $(9/2, -9/2)$ and $(-9/2, 9/2)$ states; the coupling of this $S = 0$ system to environmental degrees of freedom should be small, which means decoherence effects should also be small.

In future work, we shall use the Landau–Zener method²⁰ to determine the tunnel splitting in $[\text{Mn4}]_2$, and apply a transverse field to probe its exact influence on QTM rates (we have already confirmed that a transverse field increases the tunnelling rate, as expected for QTM). The identification of both an antiferromagnetic coupling and an exchange-bias effect in $[\text{Mn4}]_2$ demonstrates the feasibility of employing supramolecular chemistry to modulate the quantum physics of SMMs, providing a realistic method for fine-tuning the properties of these molecular nanoscale materials. This brings closer their use in devices.

Received 29 October 2001; accepted 13 February 2002.

- Sessoli, R. *et al.* High-spin molecules: $[\text{Mn}_{12}\text{O}_{12}(\text{O}_2\text{CR})_{16}(\text{H}_2\text{O})_4]$. *J. Am. Chem. Soc.* **115**, 1804–1816 (1993).
- Sessoli, R., Gatteschi, D., Caneschi, A. & Novak, M. A. Magnetic bistability in a metal-ion cluster. *Nature* **365**, 141–143 (1993).
- Christou, G., Gatteschi, D., Hendrickson, D. N. & Sessoli, R. Single-molecule magnets. *MRS Bull.* **25**, 66–71 (2000).
- Aubin, S. M. J. *et al.* Distorted $\text{Mn}^{\text{IV}}\text{Mn}^{\text{III}}$ cubane complexes as single-molecule magnets. *J. Am. Chem. Soc.* **118**, 7746–7754 (1996).
- Friedman, J. R., Sarachik, M. P., Tejada, J. & Ziolo, R. Macroscopic measurement of resonant magnetization tunnelling in high-spin molecules. *Phys. Rev. Lett.* **76**, 3830–3833 (1996).
- Thomas, L. *et al.* Macroscopic quantum tunneling of magnetization in a single crystal of nanomagnets. *Nature* **383**, 145–147 (1996).
- Wernsdorfer, W. & Sessoli, R. Quantum phase interference and parity effects in magnetic molecular clusters. *Science* **284**, 133–135 (1999).
- Leuenberger, M. N. & Loss, D. Quantum computing in molecular magnets. *Nature* **410**, 789–793 (2001).
- Hendrickson, D. N. *et al.* Photosynthetic water oxidation center: spin frustration in distorted cubane $\text{Mn}^{\text{IV}}\text{Mn}^{\text{III}}$ model complexes. *J. Am. Chem. Soc.* **114**, 2455–2471 (1992).
- Desiraju, G. R. The C–H···O hydrogen bond: structural implications and supramolecular design. *Acc. Chem. Res.* **29**, 441–449 (1996).
- Freytag, M. & Jones, P. G. Hydrogen bonds C–H···Cl as a structure-determining factor in the gold(I) complex bis(3-bromopyridine)gold(I) dichloroaurate(I). *Chem. Commun.* 277–278 (2000).
- Aullón, G., Bellamy, D., Brammer, L., Bruton, E. A. & Orpen, A. G. Metal-bound chlorine often

accepts hydrogen bonds. *Chem. Commun.* 653–654 (1998).

- Raymo, F. M., Bartberger, M. D., Houk, K. N. & Stoddart, J. F. The magnitude of [C–H···O] hydrogen bonding in molecular and supramolecular assemblies. *J. Am. Chem. Soc.* **123**, 9264–9267 (2001).
- Jeffrey, G. A. & Saenger, W. *Hydrogen Bonding in Biological Structures* (Springer, Berlin, 1991).
- Desiraju, G. R. *Crystal Engineering: the Design of Organic Solids* (Elsevier, Amsterdam, 1989).
- Jones, P. G. & Ahrens, B. Bis(diphenylphosphino)methane and related ligands as hydrogen bond donors. *Chem. Commun.* 2307–2308 (1998).
- Xu, C. *et al.* Synthesis, molecular structures and fluxional behavior of dpmm-bridged complexes of platinum(II) with linear gold(I), trigonal silver(I), or tetrahedral mercury(II) centers. *Organometallics* **15**, 3972–3979 (1996).
- Wemple, M. W., Tsai, H.-L., Folting, K., Hendrickson, D. N. & Christou, G. Distorted cubane $[\text{Mn}_4\text{O}_3\text{Cl}]^{6+}$ complexes with arenecarboxylate ligation: crystallographic, magnetochemical and spectroscopic characterization. *Inorg. Chem.* **32**, 2025–2031 (1993).
- Carlin, R. L. *Magnetochemistry* (Springer, Berlin, 1986).
- Wernsdorfer, W. Classical and quantum magnetization reversal studies in nanometer-sized particles and clusters. *Adv. Chem. Phys.* **118**, 99–190 (2001).
- Sangregorio, C. *et al.* Quantum tunnelling of the magnetization in an iron cluster nanomagnet. *Phys. Rev. Lett.* **78**, 4645–4648 (1997).
- Aubin, S. M. J. *et al.* Half-integer-spin, single-molecule magnet exhibiting resonant magnetization tunnelling. *J. Am. Chem. Soc.* **120**, 839–840 (1998).
- Aubin, S. M. J. *et al.* Resonant magnetization tunnelling in the trigonal pyramidal $\text{Mn}^{\text{IV}}\text{Mn}^{\text{III}}$ complex $[\text{Mn}_4\text{O}_3\text{Cl}(\text{O}_2\text{CMe})_5(\text{dbm})_3]$. *J. Am. Chem. Soc.* **120**, 4991–5004 (1998).
- Aubin, S. M. J. *et al.* Resonant magnetization tunnelling in the half-integer-spin single-molecular magnet $[\text{PPh}_4][\text{Mn}_{12}\text{O}_{12}(\text{O}_2\text{CEt})_{16}(\text{H}_2\text{O})_4]$. *Chem. Commun.* 803–804 (1998).
- Boskovic, C., Pink, M., Huffman, J. C., Hendrickson, D. N. & Christou, G. Single-molecule magnets: ligand-induced core distortion and multiple Jahn–Teller isomerism in $[\text{Mn}_{12}\text{O}_{12}(\text{OAc})_8(\text{O}_2\text{PPh}_2)_8(\text{H}_2\text{O})_4]$. *J. Am. Chem. Soc.* **123**, 9914–9915 (2001).
- Andres, H. *et al.* Inelastic neutron scattering and magnetic susceptibilities of the single-molecule magnets $[\text{Mn}_4\text{O}_3\text{X}(\text{OAc})_3(\text{dbm})_3]$ (X = Br, Cl, OAc, and F): variation of the anisotropy along the series. *J. Am. Chem. Soc.* **122**, 12469–12477 (2000).
- Barbara, B. & Chudnovsky, E. M. Macroscopic quantum tunneling in antiferromagnets. *Phys. Lett. A* **145**, 205–208 (1990).
- Awschalom, D. D., Smyth, J. F., Grinstein, G., DiVincenzo, D. P. & Loss, D. Macroscopic quantum tunneling in magnetic proteins. *Phys. Rev. Lett.* **68**, 3092–3095 (1992).
- Gider, S., Awschalom, D. D., Douglas, T., Mann, S. & Chaparala, M. Classical and quantum magnetic phenomena in natural and artificial ferritin proteins. *Science* **268**, 77–80 (1995).
- Tejada, J. *et al.* Does macroscopic quantum coherence occur in ferritin? *Science* **272**, 424–426 (1996).

Acknowledgements

We thank the US National Science Foundation for support. We also thank A. Benoit, D. Mailly and C. Thirion for help in the development of the micro-SQUID technique, and B. Barbara for his support.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.C. (e-mail: christou@chem.ufl.edu).

Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing

Masakazu Matsumoto, Shinji Saito & Iwao Ohmine

Chemistry Department, Nagoya University, Chikusa-ku, Nagoya, Japan 464-8602

Upon cooling, water freezes to ice. This familiar phase transition occurs widely in nature, yet unlike the freezing of simple liquids^{1–3}, it has never been successfully simulated on a computer. The difficulty lies with the fact that hydrogen bonding between individual water molecules yields a disordered three-dimensional hydrogen-bond network whose rugged and complex global potential energy surface^{4–6} permits a large number of possible network configurations. As a result, it is very challenging to reproduce the freezing of ‘real’ water into a solid with a unique crystalline structure. For systems with a limited number of possible disordered hydrogen-bond network structures, such as

confined water, it is relatively easy to locate a pathway from a liquid state to a crystalline structure^{7–9}. For pure and spatially unconfined water, however, molecular dynamics simulations of freezing are severely hampered by the large number of possible network configurations that exist. Here we present a molecular dynamics trajectory that captures the molecular processes involved in the freezing of pure water. We find that ice nucleation occurs once a sufficient number of relatively long-lived hydrogen bonds develop spontaneously at the same location to form a fairly compact initial nucleus. The initial nucleus then slowly changes shape and size until it reaches a stage that allows rapid expansion, resulting in crystallization of the entire system.

The difficulties associated with simulating the freezing of water resemble those of the protein-folding problem: real proteins fold into a unique native structure from numerous denatured states^{10,11}, but attempts to simulate this process for large proteins remain, so far, unsuccessful.

Our molecular dynamics (MD) simulation of the water-freezing process involves thermalization of pure water at a high temperature, followed by quenching (at time $t = 0$) to a low temperature, 230 K.

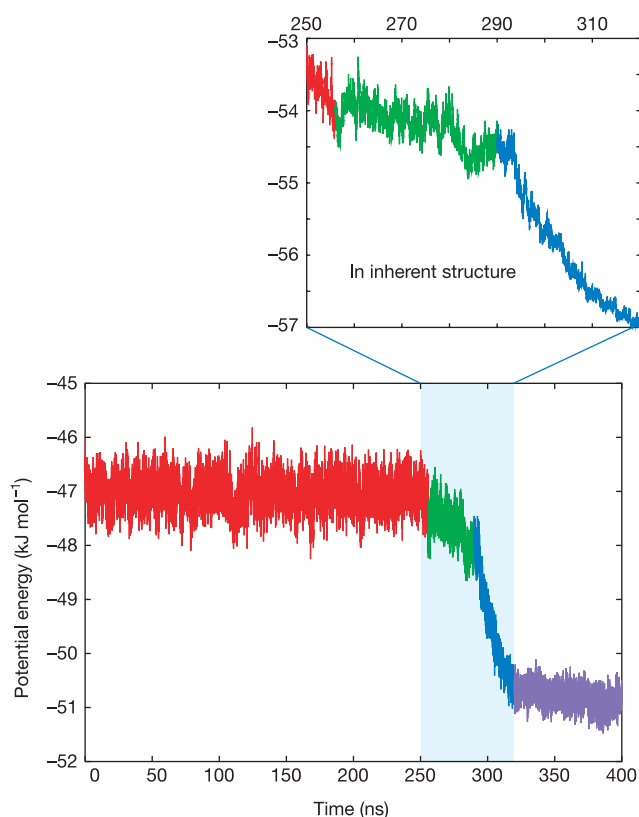


Figure 1 The total potential energies of the instantaneous structures in the trajectory for 512 water molecules after quenching (at $t = 0$ ns) from higher temperature to 230 K. Sufficient thermalization at the high temperature (400 K) is performed before the temperature jump. A constant-temperature and constant-volume molecular dynamics (MD) method is used for a system with 512 water molecules, confined in a cubic box under a periodic boundary condition. The TIP4P water model is used. The inset shows the total potential energies of the inherent structures, corresponding to the instantaneous structures in the trajectory for $t = 250$ –320 ns. Inherent structures are calculated by local quenching at 10-ps intervals. Energy is in kJ mol^{-1} and time in ns (10^{-9} s). We note that the freezing process can be divided into four stages: (1) a long quiescent period (indicated by a red line); (2) a slow energy-decreasing period (green line); (3) a fast energy-decreasing period (dark blue line); and (4) a crystallization-completion period (purple line).

The system is then monitored for crystallization for $t > 0$ while keeping the temperature at 230 K. At this temperature, water is in a supercooled state. Because the density of our system (0.96 g cm^{-3}) is slightly lower than that of normal liquid water, the degree of supercooling and hence also the nucleation rate is increased. Nevertheless, many MD trajectory calculations, each longer than $1 \mu\text{s}$, are needed to obtain a trajectory that leads to crystallization.

Figure 1 gives the total potential energy as a function of time for one such trajectory in a system of 512 water molecules. The data show that there are four stages in the freezing process: (1) a long quiescent period with relatively constant potential energy ($t = 0$ –256 ns); (2) a short period during which the potential energy slowly decreases ($t = 256$ –290 ns); (3) a short period during which the potential energy decreases rapidly ($t = 290$ –320 ns); and (4) a final period with reduced but relatively constant potential energy and during which the ice structure fully forms ($t > 320$ ns). Water in the quiescent period is in a supercooled liquid state, exhibiting intermittent collective motions and energy fluctuations associated with hydrogen bond rearrangements. The freezing process starts in stage (2). The fact that the system explores the overall relatively flat potential energy landscape for a considerable time (that is, the quiescent period) before entering the fast growing period agrees with the predictions of basic nucleation theory^{12–14}. However, the MD simulation also provides a molecular-level illustration of the water freezing process not obtainable from conventional nucleation theory¹⁴.

Figure 2 depicts the hydrogen-bond structure of the system at different times after quenching. In the quiescent period (Fig. 2a), the hydrogen-bond network contains mainly five-, six- and seven-membered rings composed of water molecules held together by hydrogen bonds. Although individual rings are destroyed and reformed continuously, the overall fraction of water molecules contained in the different ring types remains almost unchanged during the quiescent period, and is similar to the structural composition of ordinary liquid water.

Hydrogen bonds with a relatively long lifetime (τ_{life}) of more than 2 ns are identified in the figure by bright blue lines. (For comparison, the average lifetime of hydrogen bonds in liquid water is about 1 ps at $T = 300$ K and 180 ps at 230 K.) In the quiescent stage, these 'long-lasting' hydrogen bonds appear occasionally and randomly at various locations (Fig. 2a). More than 90% of water molecules forming these 'long-lasting' bonds are mostly four-coordinated and their average potential energy is lower than that of other water molecules by more than 2 kJ mol^{-1} .

At $t \approx 256$ ns, a polyhedral structure composed of long-lasting hydrogen bonds forms spontaneously (see the circled region in Fig. 2b). This polyhedron changes its position and shape by altering hydrogen bonds with surrounding water molecules, grows slowly and finally 'anchors' at the position it occupies at $t \approx 290$ ns (Fig. 2c). The nucleus then expands rapidly, by transforming its hydrogen-bond network elements into mainly six-membered rings, which percolate through the entire three-dimensional space of the system. At the same time, the system decreases its total potential energy rapidly (the dark-blue line in Fig. 1). At the end of this period of rapid growth, a 'stacked honeycomb' structure, consisting of six-membered rings, is established throughout most of the system (Fig. 2d).

The number of six-member rings (N_{6R}) fluctuates but never increases significantly before this period of fast growth occurs. We also monitored the so-called Q_6 parameter¹⁵, a commonly used and highly sensitive index for orientational order, which in our system indicates whether or not the relative orientations of hydrogen bonds at individual water sites are coherently ordered. This parameter rises only once the fast growth stage has started, indicating that neither N_{6R} nor Q_6 are order parameters suitable for describing the entire freezing process. During the final freezing stage ($t > 320$ ns), the system undergoes a very slow completion process that involves

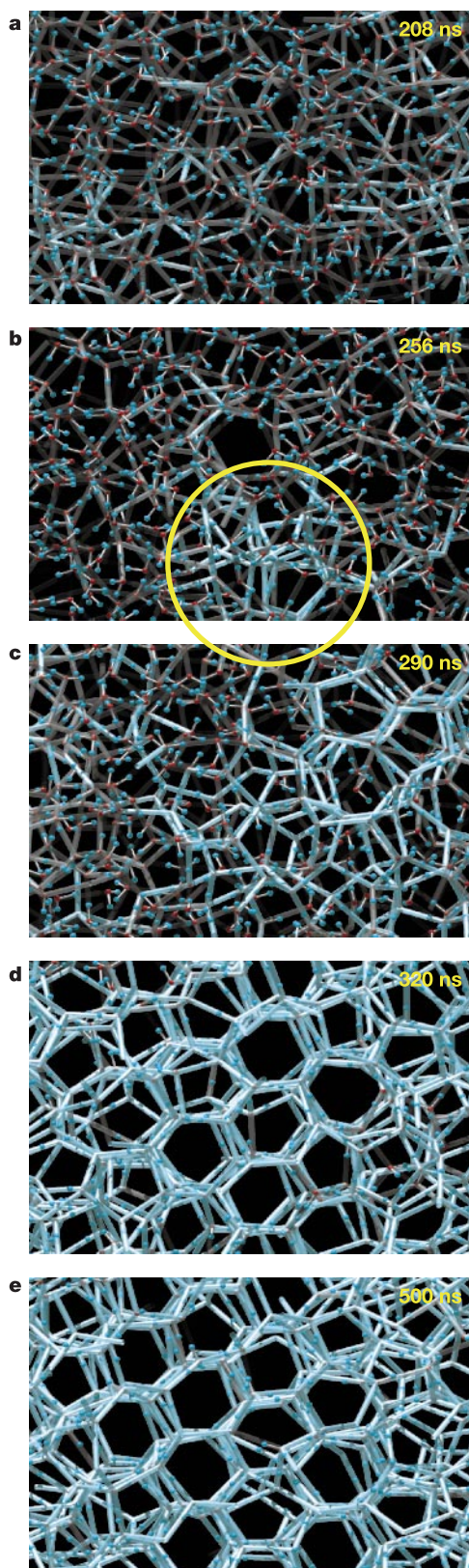


Figure 2 The hydrogen bond network structure of water at a given time in inherent structures. **a**, At time $t = 208$ ns; **b**, at $t = 256$ ns; **c**, at $t = 290$ ns; **d**, at $t = 320$ ns; and **e**, at $t = 500$ ns. Lines indicate hydrogen bonds among water molecules, and the intermolecular bonds of water participating in such hydrogen bonds. Bright blue lines indicate 'long-lasting' hydrogen bonds, which have persisted longer than a specified threshold value before a given time t . The brightest blue lines are those with a lifetime longer than 2 ns ($\tau_{\text{life}} > 2$ ns). An initial nucleus is formed in the region circled in **b**.

gradual formation of a perfect honeycomb structure (Fig. 2e) and a concomitant decrease of the potential energy (the purple line in Fig. 1).

Figure 3 illustrates the number of long-lasting hydrogen bonds ($\tau_{\text{life}} > 2$ ns) as a function of time. This number changes intermittently in the quiescent period, fluctuates in the initial nucleation period and then rapidly increases after that. During the quiescent period, most of the long-lasting bonds appear intermittently, reminiscent of the intermittent collective motion of water molecules that occurs as a result of extensive hydrogen-bond network rearrangement dynamics of liquid-state water^{4–6,16}. A Fourier transform of the fluctuation of the number of long-lasting hydrogen bonds during the quiescent period indeed yields a $1/f$ -type power spectrum with the same slope as the structural fluctuation associated with collective motions^{17–19}. Such intermittent dynamics is a characteristic of so-called frustrated systems^{20,21}, such as liquid water.

Although the long-lasting hydrogen bonds are relatively stable, they are randomly scattered and therefore usually unable to form a large stable structure; during the quiescent period, even the long-lasting bonds thus dissipate rapidly again after formation. However, at $t \approx 256$ ns several long-lasting bonds appear by chance in the same location, where they form a polyhedral structure (see Fig. 4) which evolves into a stable initial ice nucleus. The observation that the initial ice nucleus forms by chance as a result of random hydrogen-bond network rearrangement dynamics implies that the length of the quiescent period is randomly distributed.

Using a simple lattice model that captures the tetrahedral hydrogen-bond structure at individual water sites^{22,23} to simulate water freezing, we obtained many freezing trajectories. The length of the quiescent periods obtained with this model has a Poisson distribution; that is, the duration of this period is indeed randomly distributed. Similarly, MD calculations for a small system of only 64 water molecules indicated that the length of the quiescent period preceding freezing is apparently random. It should be emphasized, however, that the initial nucleus formation is not an entirely random process; that is, the initial nucleus is created as a result of

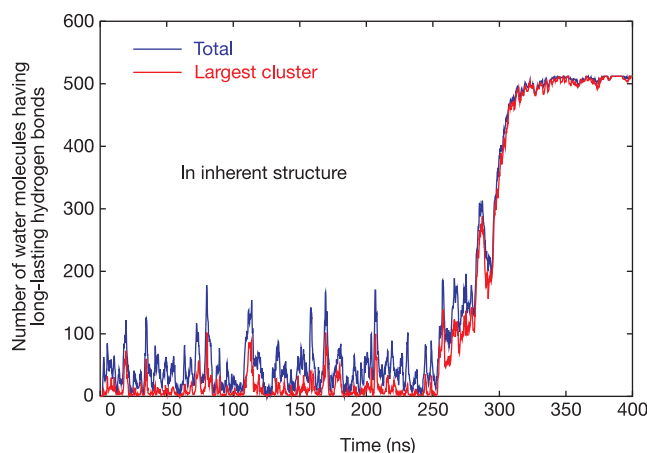


Figure 3 Number of water molecules having long-lasting hydrogen bonds ($\tau_{\text{life}} > 2$ ns) and the size of the largest cluster in inherent structures. The number of water molecules having 'long-lasting' hydrogen bonds is indicated by a dark-blue line. See Methods and Fig. 2 legend for the definition of 'long-lasting' hydrogen bonds. The red line indicates the number of water molecules belonging to the largest cluster. A cluster here is defined as a group of water molecules connected by long-lasting hydrogen bonds. We note that the number fluctuates intermittently in the quiescent period ($t = 0$ –256 ns), for example, at $t = 20, 35, 70, 80, 110, 160, 170$ and 207 ns. A Fourier transform of this number fluctuation during the quiescent period yields a $1/f$ -type power spectrum.

intermittent dynamics (strongly correlated molecular motions), with randomness arising only owing to the need for localization of 'long-lasting' hydrogen bonds.

Figure 4 illustrates how the shape (asphericity) of the nucleus changes during the freezing process when the potential energy slowly decreases ($t = 256\text{--}290$ ns in the present trajectory). The initial nucleus formed at $t = 256$ ns is a polyhedron that contains different ring types, does not belong to any well-classified ice or water structure and slightly extends in space. At $t = 260$ ns, it rapidly changes to a more compact form (that is, one of small asphericity) while decreasing in cluster size. This nucleus then grows slowly while changing its shape extensively. At $t = 270$ ns, the nucleus again adopts a very compact structure, but is now relatively large; it may in fact be beyond the size of the 'critical nucleus' that plays a central role in homogeneous nucleation theory¹⁴. Restarting from the instantaneous configurations along the present trajectory, we performed more than 100 trajectory calculations using different initial molecular velocities. Trajectories restarting from this nucleus structure at $t = 270$ ns indeed crystallize quickly, similar to the

behaviour seen in the original trajectory shown in Fig. 1. In contrast, only a very small portion, 9%, of the trajectories restarting from the configurations reached during the period $t = 240\text{--}256$ ns start to crystallize within 100 ns. The percentage of the trajectories successfully crystallized increases with the restarting time; 29% for those restarting from the configurations during $t = 256\text{--}265$ ns and 60% for those during $t = 265\text{--}270$ ns. These observations imply that the formation of a nucleus with the critical or minimum size for crystallization occurs during this period, $t = 256\text{--}270$ ns. The growth of the nucleus increases around $t = 282$ ns and becomes very rapid after $t = 290$ ns (see Fig. 3). During the period when the initial nucleus is formed ($t = 256\text{--}270$ ns) the inherent structure analysis, finding the nearest local potential energy minima along the trajectory (see Methods), yields a very rugged potential energy landscape (see Fig. 1 inset). The average potential energy of individual water molecules constituting the initial nucleus is $4\text{--}6\text{ kJ mol}^{-1}$ (4–6%) lower than that of other water molecules, with the exact value depending on the shape of the nucleus (the average potential energy is lower for more compact forms).

In the quiescent period, many long-lasting hydrogen-bond clusters appear intermittently, but their shape is much less compact than that of the cluster formed at $t = 256$ ns. For example, the polyhedron at $t = 207$ ns is larger but also significantly more extended than the polyhedron seen at $t = 256$ ns (see the inset of Fig. 4) and thus rapidly disappears.

The size of the critical nucleus must sensitively depend on its shape and on the size of system employed in MD. A larger system with 4,096 water molecules also yields extensive changes of nucleus shape (asphericity) during the period of initial nucleus formation, before entering the period of rapid growth. For much larger systems, the overall potential energy change associated with the formation of the initial nucleus and its subsequent rapid growth is fairly smooth, unlike the distinct changes seen in Fig. 1 for the system of 512 water molecules. The size of this system might be too small to deal with the detailed mechanism underlying the rapid growth.

Our MD trajectories under constant-temperature and constant-volume conditions have elucidated the molecular-level mechanism involved in water freezing, but the calculations do not account for important factors that might significantly affect the nucleation rate. For example, liquid water at low temperature exhibits large-scale density fluctuations^{24,25}, which have been interpreted to mean that there are regions of low-density liquid phase that facilitate initial nucleus formation. These low-density regions induce 'wetting' of the nucleus—that is, a matching of the density and hence the lattice constant of the surrounding liquid with that of the nucleus—thus facilitating the formation of hydrogen bonds between the two states and hence rapid crystal growth. (This effect is partly included in the present calculations by using a slightly reduced water density of 0.96 g cm^{-3} .) It will be of interest to determine whether the crystallization rate diminishes if the large scale density fluctuations in water are reduced by some means², and whether such a reduction would make it possible to obtain amorphous ice instead of crystalline ice under ambient conditions. The kinetic energy released by local crystallization might provide the excess energy needed to cross over the barriers on the rugged potential energy surface. To investigate these effects, MD simulations under constant pressure and with a detailed description of local temperature variations should be performed for very large systems. Finally, in order to obtain a more quantitative molecular-level description of nucleation, a free energy surface along an 'ordering parameter' (for example, a characteristic tetrahedral ordering parameter) needs to be constructed and evaluated. Such results would allow an assessment of the validity of the nucleation rate theory derived from thermodynamics and help to identify the intrinsic dynamical effects that control the water-freezing process. □

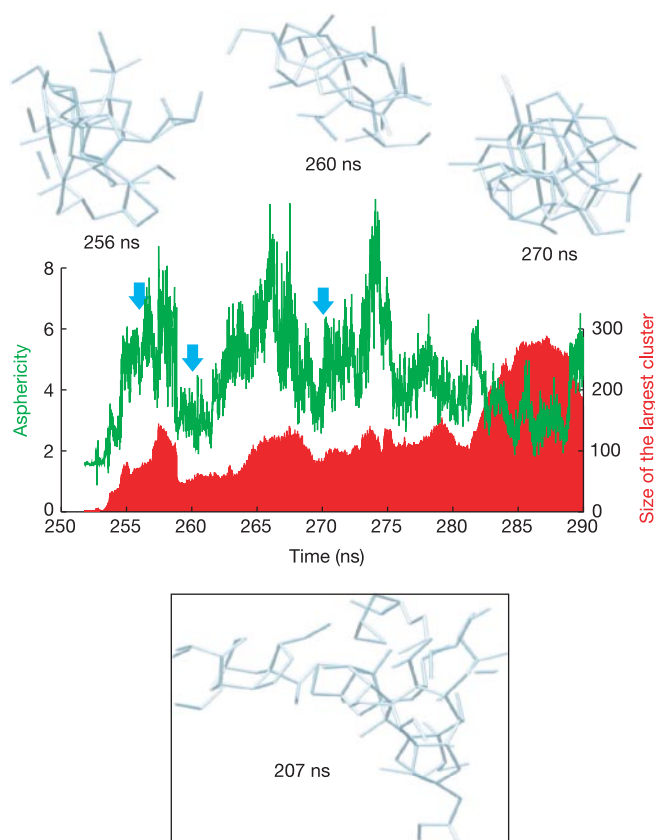


Figure 4 Asphericity and size of the largest cluster. Asphericity (green line) is calculated at 10-ps intervals for inherent structures. A cluster is defined as a group of water molecules connected by long-lasting hydrogen bonds. We assign Voronoi polyhedra to individual water molecules belonging to this nucleus. The shape of the nucleus is represented by a polyhedron which consists of these Voronoi polyhedra. Asphericity of the nucleus is defined as the cube of the surface area divided by the square of the volume of this polyhedron, renormalized to yield the asphericity of a sphere to be 1. Asphericity increases as the shape deviates from the sphere; for example, it is 1.9 for a cube. The size of the largest cluster (red bar lines) is defined as the number of water molecules belonging to the nucleus. The structures of the largest cluster (nucleus) for $t = 256$, 260 and 270 ns are also shown (blue arrows from left to right), as well as that at $t = 207$ ns (whose cluster size is 100) in the inset.

Methods

Molecular dynamics (MD) calculations are performed for 512 molecules contained in a cubic box with the periodic boundary. A time step in MD is 1 fs (femtosecond; 10^{-15} s). In the present study, a constant-volume (0.96 g cm^{-3}) and constant-temperature MD is employed. The Nose–Hoover method is used to control the temperature. An empirical water model, TIP4P, is used to describe the water–water molecular interaction. Intermolecular interaction is cut off smoothly as in previous studies^{5,6,16}. MD trajectories are performed for different temperatures, but so far only one at 230 K undergoes the crystallization for 512 water molecules. We have performed 6 trajectory calculations of the order of microseconds, but only the one that successfully crystallized is presented in this work. An individual trajectory calculation of this size system takes several months in a supercomputer. One trajectory with 220 K is found to stop in a partially crystallized structure. Those with higher temperatures have never resulted in freezing. MD calculations were also performed for various size systems containing 64, 96, 216, and 4,096 water molecules. For 4,096 water molecules, a freezing process is also attained at 230 K. Constant-pressure and constant-temperature trajectories have also been performed for various sizes of systems but crystallization has so far been achieved only for a very small system (64 water molecules). These constant-pressure, constant-temperature trajectories show that large density fluctuations are always associated with the freezing process. An inherent structure analysis²⁶ is performed by applying a local quenching method; the inherent structures are local minima of the potential wells sequentially visited by the system in a trajectory. In the inherent structure description, vibrational motions are thus removed, revealing the fundamental hydrogen-bond structure changes along the trajectory^{4–6,26}.

Received 22 December 2001; accepted 19 February 2002.

1. Swope, W. C. & Andersen, H. C. 10^6 particle molecular-dynamics study of homogeneous nucleation of crystals in a supercooled atomic liquid. *Phys. Rev. B* **41**, 7042–7054 (1990).
2. Wolde, P. R. & Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuation. *Science* **277**, 1975–1978 (1997).
3. Ball, K. D. *et al.* From topographies to dynamics on multidimensional potential energy surfaces of atomic clusters. *Science* **271**, 963–966 (1996).
4. Ohmine, I. & Saito, S. Water dynamics: fluctuation, relaxation, and chemical reactions in hydrogen bond network rearrangement. *Acc. Chem. Res.* **32**, 741–749 (1999).
5. Ohmine, I. Liquid water dynamics: collective motions, fluctuation and relaxation. *J. Phys. Chem.* **99**, 6767–6776 (1995).
6. Ohmine, I. & Tanaka, H. Fluctuation, relaxation and hydration in liquid water. Hydrogen-bond rearrangement dynamics. *Chem. Rev.* **93**, 2545–2566 (1993).
7. Svishchev, I. M. & Kusalik, P. G. Electrofreezing of liquid water: A microscopic perspective. *J. Am. Chem. Soc.* **118**, 649–654 (1996).
8. Nada, N. & Furukawa, Y. Anisotropic growth kinetics of ice crystals from water studied by molecular dynamics simulation. *J. Cryst. Growth* **169**, 587–597 (1996).
9. Koga, K., Tanaka, H. & Zeng, X. C. First-order transition in confined water between high-density liquid and low-density amorphous phases. *Nature* **408**, 564–567 (2000).
10. Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
11. Dobson, C. M. & Karplus, M. The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92–101 (1999).
12. Kramer, B. *et al.* Homogeneous nucleation rates of supercooled water measured in single levitated microdroplets. *J. Chem. Phys.* **111**, 6521–6527 (1999).
13. Bartell, L. S. Nucleation rates in freezing and solid-state transitions. *J. Phys. Chem.* **99**, 1080–1087 (1995).
14. Abraham, F. F. *Homogeneous Nucleation Theory* (Academic, New York, 1974).
15. Steinhart, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784–805 (1983).
16. Ohmine, I., Tanaka, H. & Wolynes, P. G. Large local energy fluctuation in water. *J. Chem. Phys.* **89**, 5852–5860 (1988).
17. Sasai, M., Ramaswamy, R. & Ohmine, I. Long time fluctuation of liquid water. *J. Chem. Phys.* **96**, 3045–3053 (1992).
18. Walrafen, G. E., Hokmabadi, M. S., Yang, W.-H., Chu, Y. C. & Monosmith, B. Collision-induced Raman scattering from water and aqueous solutions. *J. Phys. Chem.* **93**, 2909–2917 (1989).
19. Saito, S. & Ohmine, I. Translational and orientational dynamics of a water cluster (H_2O)₁₀₈ and liquid water. *J. Chem. Phys.* **102**, 3566–3576 (1995).
20. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1–77 (2000).
21. Weissman, M. B. $1/f$ noise and other slow, non-exponential kinetics in condensed matter. *Rev. Mod. Phys.* **60**, 537–571 (1988).
22. Sasai, M. Instabilities of hydrogen bond network in liquid water. *J. Chem. Phys.* **93**, 7329–7341 (1990).
23. Matsumoto, M. & Ohmine, I. New approach to dynamics of hydrogen bond network in liquid water. *J. Chem. Phys.* **104**, 2705–2712 (1996).
24. Mishima, O. & Stanley, H. E. The relationship between liquid, supercooled and glassy water. *Nature* **396**, 329–335 (1998).
25. Sastry, S., Debenedetti, P. G., Sciortino, F. & Stanley, H. E. Singularity-free interpretation of the thermodynamics of supercooled water. *Phys. Rev. E* **53**, 6144–6154 (1996).
26. Stillinger, F. H. & Weber, T. A. Packing structures and transitions in liquids and solids. *Science* **225**, 983–989 (1984).

Acknowledgements

We thank J. Jortner, and S. Sastry for many stimulating discussions and encouragement. We also thank H. Tanaka, A. Baba and H. Inagaki for cooperative work and valuable discussions. The present study is partially supported by the Grant-in-Aid Scientific Research on Priority Area of 'Condensed Phase Chemical Reaction Dynamics', and by the

Grant-in-Aid for Scientific Research. Calculations were carried out at the Nagoya University Computation Center and the Research Center for Computational Science.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.O. (e-mail: ohmine@aquac.chem.nagoya-u.ac.jp).

Active fluidization of polymer networks through molecular motors

D. Humphrey*, C. Duggan*, D. Saha*, D. Smith* & J. Käs*†‡§

* Center for Nonlinear Dynamics, † Institute for Cellular and Molecular Biology, and ‡ Texas Materials Institute, and § Center for Nano and Molecular Science, University of Texas at Austin, Texas 78712, USA

Entangled polymer solutions and melts exhibit elastic, solid-like resistance to quick deformations and a viscous, fluid-like response to slow deformations. This viscoelastic behaviour reflects the dynamics of individual polymer chains driven by brownian motion¹: since individual chains can only move in a snake-like fashion through the mesh of surrounding polymer molecules, their diffusive transport, described by reptation^{2–4}, is so slow that the relaxation of suddenly imposed stress is delayed. Entangled polymer solutions and melts therefore elastically resist deforming motions that occur faster than the stress relaxation time. Here we show that the protein myosin II permits active control over the viscoelastic behaviour of actin filament solutions. We find that when each actin filament in a polymerized actin solution interacts with at least one myosin minifilament, the stress relaxation time of the polymer solution is significantly shortened. We attribute this effect to myosin's action as a 'molecular motor', which allows it to interact with randomly oriented actin filaments and push them through the solution, thus enhancing longitudinal filament motion. By superseding reptation with sliding motion, the molecular motors thus overcome a fundamental principle of complex fluids: that only depolymerization makes an entangled, isotropic polymer solution fluid for quick deformations.

Actin and myosin II are the key elements of the contractile machinery in muscle cells. Myosin's motor domains, or 'heads', bind actin filaments (F-actin) and exploit adenosine triphosphate (ATP) hydrolysis to generate a force to move along polar actin filaments towards the positive end. *In vitro* and in non-muscle cells, myosin II proteins assemble into multimeric bipolar structures with motor domains on both ends^{6–9} known as minifilaments. Several of the myosin heads at the end of these minifilaments attach to actin filaments. When ADP (adenosine diphosphate) is substituted for ATP in solution, the minifilaments crosslink F-actin via the inactive heads. In this case the actin–myosin sample assumes a solid, gelatin-like consistency (Fig. 1a). In the presence of ATP, myosin is active and the sample behaves like a fluid (Fig. 1a). The myosin heads push the actin filaments and collectively induce sliding of filaments in the presence of ATP¹⁰ (Fig. 1b). Similar to two-dimensional motility assays¹¹ in which F-actin moves on myosin heads adsorbed to a cover slide, filaments are unidirectionally pushed by myosin towards their negative ends. The sliding filaments are oriented at a variety of angles with respect to each other.

By inducing active filament sliding instead of thermally driven

filament transport, myosin II has a large effect on the dynamics of actin filaments in networks. Rheological measurements of the frequency-dependent (that is, stress-rate dependent) complex shear modulus, G^* , characterize the effect of these changes in polymer dynamics on actin's viscoelastic properties (Fig. 2). The elastic part of G^* is referred to as the storage modulus G' , while the viscous part is known as the loss modulus G'' . The relation between G' and G'' describes to what extent the polymer sample responds elastically¹². At low frequencies, the polymer solution behaves like a fluid and responds viscously ($G' \leq G''$), whereas at higher frequencies, the sample responds like rubber by resisting elastically

($G' > G''$). Thus, frequency-dependent rheology is ideally suited to distinguish elastic from fluid solutions.

We compared entangled solutions of actin filaments to isotropic actin networks with active or inactive myosin II dispersed in them. An actin monomer concentration of $36 \mu\text{M}$ and myosin concentration of $0.14 \mu\text{M}$ in the experiments resulted in a ratio of myosin proteins to actin filaments of ~ 15 , with 2–16 myosin proteins assembled to small bipolar minifilaments (see light-scattering data in Supplementary Information). The addition of inactive myosin led to an increase in the plateau storage modulus G' of a factor of 6 ± 1.3 (see Fig. 2). Inactive myosin acts as a crosslinker, and this gelation process increases the elastic strength of the sample¹³. The addition of ATP, which induces filament sliding, caused a decrease of G' to $40 \pm 10\%$ of G' for actin solutions without myosin (see Fig. 2). Thus, the elastic strength dropped below the value for a plain entangled actin solution. Moreover, G' approximately equalled G'' , indicating that active myosin renders the network more liquid-like.

The interactions of single chains with surrounding polymers, which underlie the viscoelastic behaviour, can be described by an effective-medium treatment often visualized by a 'tube' arising from the topological constraints of neighbouring polymer chains¹. As these constraints oppose deformations¹⁵, stress is relaxed when filaments have left their original, confining tube (independent of the direction the filament moves with respect to the deforming stress). To allow visualization of the mechanism that enables myosin to modulate the elastic strength of actin networks, rhodamine-phalloidin-labelled actin filaments and myosin were incorporated in networks of unlabelled actin filaments. Fluorescence microscopy could then be used to follow the dynamics of individual actin chains within the network⁴. A direct comparison of single filament motion between a network with active motors and an entangled solution of actin filaments shows large differences in speed and directionality (Fig. 3). In the absence of motors, the filament undergoes random brownian motion confined by the surrounding unlabelled actin

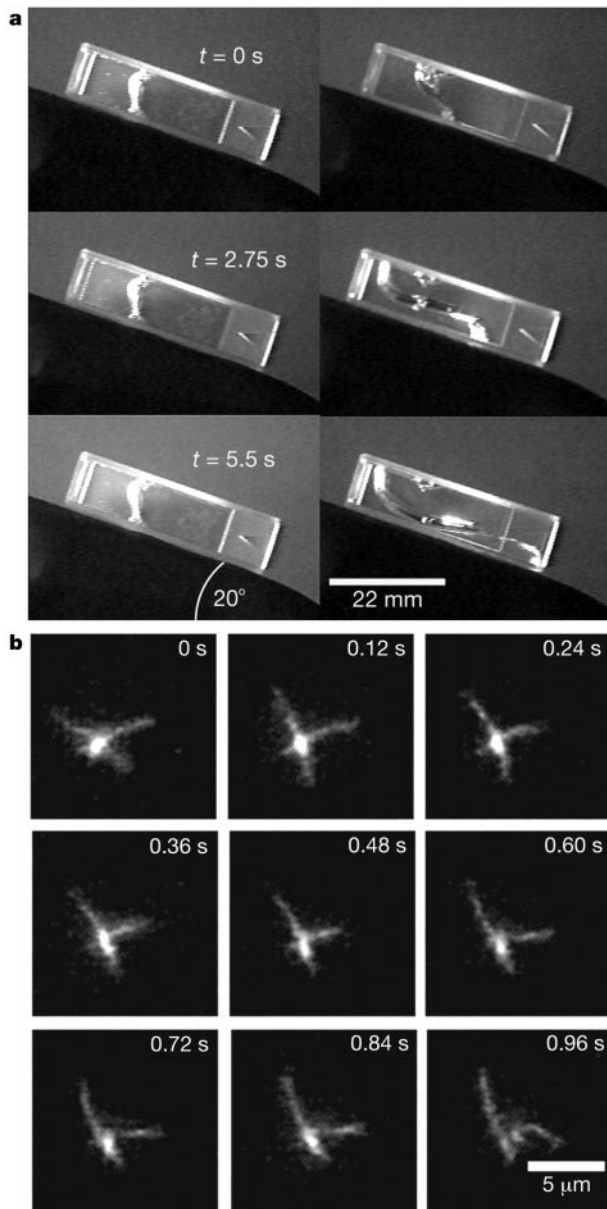


Figure 1 Macroscopic and microscopic influence of bipolar myosin minifilaments on actin networks. **a**, Comparison of the flow properties of actin networks with inactive (left side) or active (right side) myosin dispersed in them. For this purpose a cuvette filled to one-third with sample was tilted from its initial vertical position. With ADP and inactive myosin the samples gelled and behaved like an elastic solid. In the presence of ATP or when caged ATP was released, the active motors caused fluid-like flow properties. **b**, In the presence of ATP, two rhodamine-phalloidin-labelled actin filaments in dilute solution slide along each other while connected by BODIPY FL-labelled myosin (bright spot). The left filament moves to the upper left side with respect to the myosin minifilament and the right filament moves to the right. In the last picture filaments disengage.

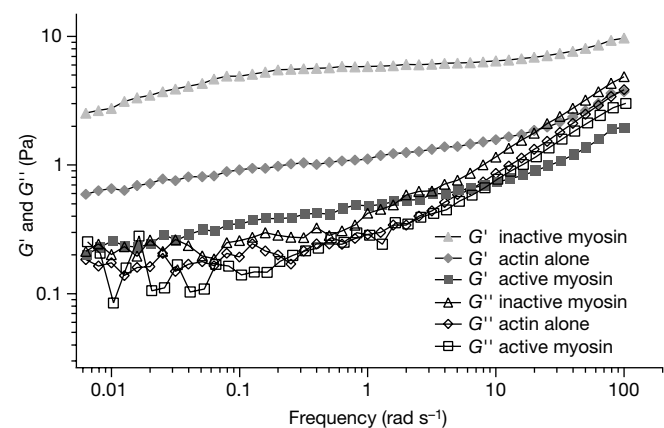


Figure 2 Elastic strength of actin and actin-myosin networks ($36 \mu\text{M}$ actin, $0.14 \mu\text{M}$ myosin, $500 \mu\text{M}$ ATP/ADP). The filled markers display the storage modulus, G' , a measure of the elastic strength. The open markers show the loss modulus, G'' , which characterizes the viscous component. The inactive myosin-actin network shows the highest value of G' . When myosin was activated by ATP the value of G' decreased below the value for the entangled actin solution. Below 0.03 rad s^{-1} G' and G'' are comparable signifying, in effect, a fluid solution. According to our microscopic stress relaxation experiments (see Fig. 4 inset) the onset of liquefaction should occur at 0.1 rad s^{-1} . The slightly lower onset is caused by the polydispersity of the sample. Between 0.03 and 5 rad s^{-1} a small elastic plateau remains for the active actin-myosin network. At higher frequencies ($> 5 \text{ rad s}^{-1}$) the rheometer probes the internal dynamics of filaments indicated by the typical crossover between G' and G'' . In the high-frequency regime the active motors cause a transition from a coil-like to a rod-like behaviour¹⁶ inducing the early onset of the internal dynamics regime.

filaments, with no net transport being detectable on a timescale of ten seconds. The filament remains enclosed in the original tube formed by the surrounding filaments. In actin–myosin networks, myosin induces longitudinal sliding of filaments when ATP is added. At least 98% of the studied filaments in these dense three-dimensional entangled F-actin networks displayed sliding activity when active myosin was present, with an experimentally determined average sliding speed of $1.3 \pm 0.3 \mu\text{m s}^{-1}$. Under ADP conditions, filaments are immobilized owing to crosslinking by inactive myosin (data not shown).

In actin–myosin solutions under ATP conditions, the time for a filament to disengage from its original tube confinement, τ , dras-

tically decreases, because filament sliding prevails over random backward and forward motions (Fig. 4 inset). For deformations slower than τ , solutions of actin filaments appear fluid, but for quicker motions an elastic stress opposes the distortion because filaments are confined in their surrounding tubes. This stress is relaxed when a filament leaves its original tube, that is, moves in a random direction over a distance equal to its length. As can be seen in the inset of Fig. 4, the tube disengagement time (that is, local stress relaxation time), τ , depended linearly on the filament length, L , when motors caused filament sliding. In entangled actin solutions the relaxation time depended on the filament length as $\tau \propto L^3$, in agreement with the reptation concept^{2,4}. Thus, in actin networks with a polydisperse filament length distribution ($\sim 1\text{--}130 \mu\text{m}$) and a mean filament length of $10 \mu\text{m}$, the average stress relaxation time significantly decreased from $500 \pm 100 \text{ s}$ for pure actin solutions to $8 \pm 0.4 \text{ s}$ when active myosin was added. Inactive myosin froze filament transport by crosslinking the filaments; no relaxation occurred over an observation time of 20 minutes.

As a result, an actin solution with active myosin will mostly relieve macroscopic stress within $\sim 8 \text{ s}$ and should be fluid-like for all deformations performed at a rate slower than $1/8 \text{ s}^{-1}$, enabling the sample to flow out of a cuvette (see Fig. 1a). Consequently, the rheological behaviour should change to $G' \approx G''$ for shear frequencies below $\sim 0.1 \text{ rad s}^{-1}$. Nevertheless, a remnant indication of elasticity was found even below $\sim 0.1 \text{ rad s}^{-1}$ in networks of actin filaments and active myosin (Fig. 2). Because this decrease in residual elasticity directly correlated with our efforts to purify the myosin from dysfunctional myosin heads without motor activity (also known as dead myosin), we conclude that this minor elastic

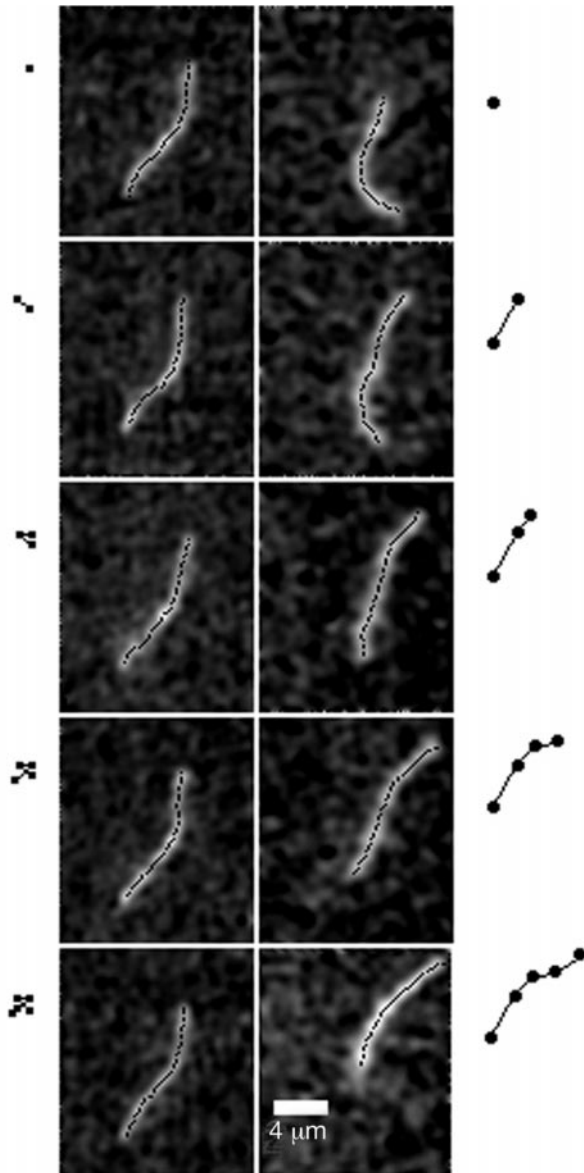


Figure 3 Motion of a single actin filament in an F-actin solution or an actin–myosin network. Fluorescently labelled filaments are embedded in an unlabelled matrix of filaments (left side) or in an unlabelled network of filaments with active myosin minifilaments present (right side). The time interval between successive pictures is 2.5 s. The left filament solely undergoes brownian motion restricted by the surrounding filaments. Within 10 s no net transport of the polymer chain was observed. The right filament shows directional sliding induced by myosin, resulting in a quick transport of the filament. The dashed black lines represent the digital traces of filament contours used to analyse filament motion. The two outer columns depict the motion of the upper filament ends without (squares) and with (circles) myosin.

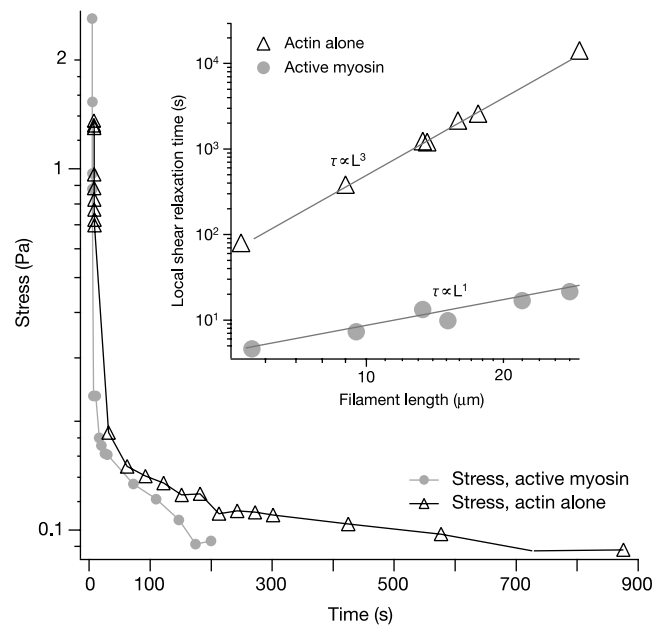


Figure 4 Temporal stress relaxation behaviour in actin and actin–myosin networks. The main plot displays the bulk stress relaxation behaviour in response to a prior 3% step strain. The relaxation of an entangled F-actin solution (triangles) is compared to actin networks with active myosin (filled circles). Stress below 0.1 Pa cannot be reliably detected with our rheometer. The inset shows a logarithmic plot of the time, τ , a filament needs to move a distance equal to its length (that is, the local stress relaxation time) either in a solution of actin filaments (triangles) or in an actin–myosin network under ATP conditions (filled circles) as function of actin filament length, L . Filaments differing in their length by less than $1 \mu\text{m}$ have been combined into one data point. The relative errors in measurements of τ are 19% and 5% for reptating and sliding filaments, respectively. In an entangled solution, $\tau \propto L^3$; this changes to a linear dependence in the presence of active myosin motors. Furthermore, the observed τ decreases by more than an order of magnitude for active myosin.

contribution primarily originated from an unavoidable trace amount of dead myosin crosslinking the filaments (see Supplementary Information). A density of 1 filament pair crosslinked by dead myosin in 100,000 pairs is sufficient to cause the remaining elastic response^{14,15}. In other words, a small number of elastically active strands were mixed with a large number of filaments exhibiting fluid-like behaviour. The myosin-induced fluidification (reduction of relaxation time and elastic strength) shown experimentally here can be modelled by a diffusive directed motion of the polar filaments in their tube owing to increased longitudinal fluctuations along the filament contour, giving rise to a higher effective temperature¹⁶.

The change in the elastic response of active actin–myosin networks with respect to F-actin solutions is reflected in macroscopic stress relaxation experiments, in which a 3% step strain is applied and the time course of the resulting stress is monitored (see Fig. 4). Active myosin-induced filament sliding increased the macroscopic stress relaxation rate compared to an entangled F-actin solution by nearly an order of magnitude. However, a direct comparison between local stress relaxation by filament sliding and macroscopic stress relaxation is hampered by the presence of dead myosin heads and the polydispersity of F-actin. Polydispersity gives rise to different tube disengagement times contributing to the relaxation process. Moreover, some filaments are not pushed by minifilaments, while trace amounts of dead myosin heads act as crosslinks that prevent complete stress dissipation. The residual undissipated stress due to trace amounts of dead myosin is almost below the stress detection limit for the experiments summarized in Fig. 4, but was more pronounced when more dead myosin was contaminating the sample.

The typical ratios of myosin to actin filaments used in our experiments fall within the range typically found in eukaryotic cells ($n = 3.4\text{--}63$; ref. 17). Although a range of accessory proteins that induce crosslinks hinder filament sliding *in vivo*, crosslinking only holds back transiently the sliding filaments¹³. Thus the interplay between filament sliding and crosslinking may be rather complex. As a result, myosin's ability to modulate viscoelasticity may also be relevant to motile eukaryotic cells, given that the fluidization of the actin cytoskeleton is an essential step in cell motility⁵. □

Methods

Protein purification

Actin powder from rabbit skeletal muscle was prepared according to the original method of ref. 18, as modified in ref. 19. Monomeric actin (G-actin) was extracted by the method of ref. 20. Myosin II was purified from rabbit skeletal muscle according to ref. 21. To remove inactive myosin, 6.3 μM G-actin and 5.2 μM myosin were polymerized for two hours at room temperature in F buffer (10 mM Tris–HCl (pH 7.5), 2 mM MgCl_2 , 150 mM KCl, 0.2 mM CaCl_2 , 0.5 mM ATP). After two hours, 270 μl of 0.25 M ATP were added to a final ATP concentration of 5 mM. The solution was spun at 4°C at 100,000g for 90 min. The supernatant containing myosin with active molecular motors was transferred, and the concentration determined using the Bradford assay. The presence of full myosin II in the supernatant was established by SDS–PAGE and by light scattering (see Supplementary Information). This protocol is conventionally used to isolate active cleaved myosin heads, but also turned out to be efficient for full myosin. BODIPY-FL (Molecular Probes) was used to label myosin II according to the Molecular Probes protocol. For labelled and unlabelled myosin, motor activity was checked by actin filament sliding (that is, motility assay) as well as by ATPase activity monitored by a luciferase-based assay (see also Supplementary Information).

Polymer dynamics and rheology

The dynamics of individual rhodamine–phalloidin-labelled actin filaments in F-actin or in actin–myosin networks were visualized and analysed by fluorescence microscopy as previously described²². In all experiments, unless otherwise stated, labelled actin filaments were mixed with unlabelled filaments (ratio of 1:4000) in F buffer. All experiments were performed with 0.5 mM ATP or ADP present, and either with or without 140 nM myosin II at a final actin concentration of 36 μM . The unlabelled filaments were polymerized around the fluorescently labelled filaments. A 30 min polymerization time was sufficient for this concentration of actin. Based on an average filament length of 10 μm and on the fact that 14 actin monomers form an actin filament segment of 37 nm length, the ratio of myosin to actin filament n was calculated as $n = (c_{\text{myosin}}N_{\text{fil}})/c_{\text{actin}}$, where c_{actin} is the molar actin concentration, c_{myosin} is the molar myosin concentration, and N_{fil} is the average

number of monomers per filament. Although it was a rare event to find two labelled actin filaments simultaneously connected by a labelled myosin minifilament ($n = 5$, for example, Fig. 1b), good data ($n = 50$, for example, Fig. 3) have been obtained from observing a sliding labelled filament with unlabelled partners. The time a filament needs to move a distance equal to its length in the surrounding actin matrix has been measured as previously described^{4,22} for reptating filaments and determined by direct observation for the sliding filaments.

Rheological measurements were performed with a commercial fluid rheometer (Rheometrics Model 8400) and a newly developed Mooney–Couette torsion pendulum rheometer that had been designed for simultaneous microscopy of actin filaments and measurement of G^* . The rheological data were reproducible for six different actin and myosin preparations; at least three measurements were performed for each condition and protein preparation. To obtain consistent data, it is absolutely necessary to avoid large amounts of dead myosin heads with the precautions described above. Simultaneous rheology and microscopy excluded changes in the actin network structure, such as filament bundling and filament breakage, as the cause of the loss of elasticity in the presence of active myosin. Furthermore, the actin filaments maintained their random orientation in active actin–myosin networks. No myosin-induced ordering of actin filaments—as previously described for microtubules in the presence of molecular motors²³—was observed at the actin and myosin concentrations used. The presence of ATP throughout the experiments was verified by the observation of sliding filaments as well as by luciferase activity.

Received 4 February 2001; accepted 18 February 2002.

- Doi, M. & Edwards, S. F. *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
- de Gennes, P. G. Reptation of a polymer chain in the presence of fixed obstacles. *J. Chem. Phys.* **55**, 572–579 (1971).
- Perkins, T. T., Smith, D. E. & Chu, S. Direct observation of tube-like motion of a single polymer chain. *Science* **264**, 819–822 (1994).
- Käs, J., Strey, H. & Sackmann, E. Direct visualization of reptation for semiflexible actin filaments. *Nature* **368**, 226–229 (1994).
- Taylor, D. L. & Condeelis, J. S. Cytoplasmic structure and contractility in amoeboid cells. *Int. Rev. Cytol.* **56**, 57–144 (1979).
- Reisler, E., Smith, C. & Seegan, G. Myosin minifilaments. *J. Mol. Biol.* **143**, 129–145 (1980).
- Niederman, R. & Pollard, T. D. Human platelet myosin. *J. Cell Biol.* **67**, 72–92 (1975).
- Sinard, J. H. & Pollard, T. D. Acanthamoeba myosin II minifilaments assemble on a millisecond time scale with rate constants greater than those expected for a diffusion limited reaction. *J. Biol. Chem.* **265**, 3654–3660 (1990).
- Verkhovskiy, A. B., Svitkina, T. M. & Borisov, G. G. in *Cell Behaviour: Control and Mechanism of Motility* (eds Lackie, J. M., Dunn, G. A. & Jones, G. E.) 207–222 (Biochem. Soc. Symp. 65, Portland Press, London, 1999).
- Honda, H., Nagashima, H. & Asakura, S. Directional movement of F-actin in vitro. *J. Mol. Biol.* **191**, 131–133 (1986).
- Kron, S. J. & Spudis, J. A. Fluorescent actin filaments move on myosin fixed to a glass surface. *Proc. Natl Acad. Sci. USA* **83**, 6272–6276 (1986).
- Gupta, R. K. *Polymer and Composite Rheology*, 2nd edn (Marcel Dekker, New York, 2000).
- Wachsstock, D. H., Schwarz, W. H. & Pollard, T. D. Cross-linker dynamics determine the mechanical properties of actin gels. *Biophys. J.* **66**, 801–809 (1994).
- Morse, D. C. Viscoelasticity of concentrated isotropic solutions of semiflexible polymers. 2. Linear response. *Macromolecules* **31**, 7044–7067 (1998).
- Isambert, H. & Maggs, A. C. Dynamics and rheology of actin solutions. *Macromolecules* **29**, 1036–1040 (1996).
- Liverpool, T. B., Maggs, A. C. & Ajdari, A. Viscoelasticity of solutions of motile polymers. *Phys. Rev. Lett.* **86**, 4171–4174 (2001).
- Bray, D. *Cell Movements* (Garland, New York, 1992).
- Straub, F. B. Actin II. *Studies Inst. Med. Chem. Univ. Szeged* **3**, 23–27 (1943).
- Adelstein, R. S., Godfrey, J. E. & Kielly, W. W. G-actin: preparation by gel filtration and evidence for double stranded structure. *Biochem. Biophys. Res. Commun.* **12**, 34–38 (1963).
- MacLean-Fletcher, S. D. & Pollard, T. D. Viscometric analysis of the gelation of Acanthamoeba extracts and purification of two gelation factors. *J. Cell Biol.* **85**, 414–428 (1980).
- Margossian, S. S. & Lowey, S. Preparation of myosin and its subfragments from rabbit skeletal muscle. *Methods Enzymol.* **85**, 55–71 (1982).
- Käs, J. *et al.* F-Actin, a model polymer for semiflexible chains in dilute, semidilute, and liquid crystalline solutions. *Biophys. J.* **70**, 609–625 (1996).
- Nedelec, F. J., Surrey, T., Maggs, A. C. & Leibler, S. Self-organization of microtubules and motors. *Nature* **389**, 305–308 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank M.E. Chenevert for providing a commercial fluid rheometer, and J. Prost for help with the interpretation of the results. We also thank the following for discussions; H. Swinney, C. Moncman, F. Jülicher, A. Maggs, T. Liverpool, D. Morse, J. Guck, D. Martin, P.A. Janmey, F. Brochard, K. Browning and P.G. de Gennes.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.K. (e-mail: kas@chaos.ph.utexas.edu).

Mantle wedge control on back-arc crustal accretion

Fernando Martinez* & Brian Taylor†

* Hawaii Institute of Geophysics and Planetology, † Department of Geology and Geophysics, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, Hawaii 96822, USA

At mid-ocean ridges, plate separation leads to upward advection and pressure-release partial melting of fertile mantle material; the melt is then extracted to the spreading centre and the residual depleted mantle flows horizontally away¹. In back-arc basins, the subducting slab is an important control on the pattern of mantle advection and melt extraction, as well as on compositional and fluid gradients². Modelling studies³ predict significant mantle wedge effects on back-arc spreading processes. Here we show that various spreading centres in the Lau back-arc basin exhibit enhanced, diminished or normal magma supply, which correlates with distance from the arc volcanic front but not with spreading rate. To explain this correlation we propose that depleted upper-mantle material, generated by melt extraction in the mantle wedge, is overturned and re-introduced beneath the back-arc basin by subduction-induced corner flow. The spreading centres experience enhanced melt delivery near the volcanic front, diminished melting within the overturned depleted mantle farther from the corner and normal melting conditions in undepleted mantle farther away. Our model explains fundamental differences in crustal accretion variables between back-arc and mid-ocean settings.

The simple geometry of the subducted slab beneath the Lau basin⁴ (Fig. 1) and well-constrained spreading-centre kinematics⁵, together with our bathymetry and gravity compilation and previous geochemical^{6–8} and seismic^{9–11} studies, allow a comprehensive study of back-arc crustal accretion variables. We focus on the central and east Lau spreading centres (CLSC and ELSC, respectively) and the southern segments of the latter, termed the Valu Fa ridge (VFR). We refer to them here as the Lau spreading centres, although other spreading axes and rifts have been identified in the northern basin^{5,12} (Fig. 1). Figure 2 summarizes the characteristics of the Lau spreading centres, including their cross-sectional area or ‘inflation’, which is a measure of their magmatic robustness¹³ (see Supplementary Information). Basin-wide gridded gravity and bathymetry data were used to calculate the mantle Bouguer anomaly, and to remove the across-basin gradient due to slab and subduction effects (see Supplementary Information). Unlike the situation at mid-ocean ridges, this adjusted mantle Bouguer anomaly forms relative axial maxima¹⁴. The origin and magnitude of the relative maxima of this adjusted anomaly correlates with other parameters (depth, morphology, crustal structure and composition) that define distinct zones of crustal accretion (Fig. 2). We describe each zone in turn before presenting a model to explain the observations.

Zone 1 in Fig. 2 has been interpreted¹⁵ as a rifted and volcanic terrain representing crust that pre-dated the establishment of the present spreading systems. As the basin opened, the ridge system that evolved into the VFR and ELSC originated in the northern basin near the arc volcanic front and propagated asymmetrically southward into this terrain, preferentially rifted into the crust near the volcanic front, as occurs at other back-arc basins^{16,17}. Crustal accretion on these axes, however, was roughly symmetric so that with time the spreading centres separated from the volcanic front.

Located at the southern end of the spreading system and closest (40–60 km) to the volcanic front, the VFR has a narrow triangular

axial high that is more peaked than typical mid-ocean-ridge axial highs¹⁸. This shape has been attributed to the more viscous nature—compared to mid-ocean-ridge basalt (MORB)—of the basaltic andesite to rhyodacitic composition of the erupted lavas¹⁸, which have affinities to the proximal volcanic arc^{7,8}. The VFR is the shallowest part of the spreading axes, and has high inflation (1–6 km²) and low values of adjusted mantle Bouguer anomaly (Fig. 2). An axial magma chamber reflector has been imaged almost continuously south of 20° 30′ S (ref. 9), and active hydrothermal vent fields¹⁹ occur south of 21° 50′ S. Seismic refraction lines^{10,11} indicate that zone 2 crustal thicknesses are 7.5–9 km, significantly thicker than at typical mid-ocean ridges, both near the VFR axis at 22°–22° 30′ S and west of the CLSC at 18° 25′ S. Zone 2 represents crust that was generated when the spreading centre was near the volcanic front and in a magmatically robust stage, analogous to the present-day VFR. This terrain displays abyssal-hill-like morphology in a few places, but much more often shows a hummocky texture with numerous crescent-shaped highs that are concave toward the axis in plan view. This landform is characteristic of enhanced ridge volcanism forming small split volcanoes²⁰, and its prevalence throughout zone 2 implies a terrain dominated by volcanic extrusion. Although the sampled portions of the VFR show strong arc affinities^{7,8}, drilling in a small basin (ODP Site 836) and dredging of a nearby high show both arc- and MORB-like compositions in close proximity within this terrain¹⁵. Zone 2 sea floor away from the propagation boundary with zone 1 maintains shallow depths off-axis, displays low adjusted mantle Bouguer anomalies (Fig. 2), and has 7.5–9-km-thick crust^{10,11}. All of these characteristics evince a robust magmatic state on the VFR and within zone 2, even though the VFR has the lowest spreading rates (~40–60 mm yr⁻¹) of the Lau spreading centres (Fig. 2c).

The ELSC north of the VFR deepens from ~2,000 to ~3,000 m

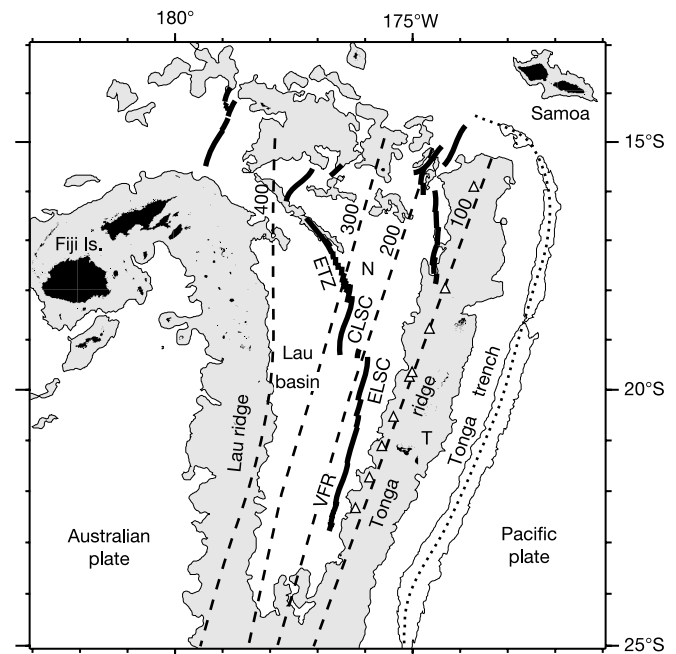


Figure 1 Location map of the Lau basin showing the back-arc spreading centres (heavy lines), trench axis (dotted line) and contours of the subducted slab (dashed lines) labelled in km. The area inside the 2,000-m bathymetry contour is shaded grey, and highlights the Lau and Tonga ridges. The 7,000-m contour surrounds the axis of the Tonga trench. Arc volcanoes are shown as open triangles. Beneath the Lau spreading centres the slab is approximately planar, strikes parallel to the volcanic front, and dips at about 45°. N, Niuafo'ou plate; T, Tonga plate; VFR, Valu Fa ridge; ELSC, east Lau spreading centre; CLSC, central Lau spreading centre; ETZ, extensional transform zone.

and develops a negative cross-sectional area (axial region deeper than flanks) as spreading rates increase from ~ 60 to ~ 95 mm yr $^{-1}$ and separation from the volcanic front increases from 60 to 110 km (Fig. 2). Adjusted mantle Bouguer anomalies increase northward by >30 mGal from the relative axial minimum at the 22° 13' S overlapping spreading centres on the VFR, indicating thinning crust and/or increasing density. Some northward increase in crustal density is likely as compositions change from typical andesites on the VFR to basalts on the ELSC^{6,7}, but isostatic calculations indicate that a very large density increase (for example, from 2,550 to 2,880 kg m $^{-3}$) would be required to achieve this deepening without some crustal thinning. As discussed below, this is unlikely given the porosity and Fe content of northern ELSC lavas⁶. Despite fast spreading rates, an axial and across-axis seismic survey⁹ did not

image a magma chamber reflector north of 20° 30' S (Fig. 2). Geochemical analysis of rocks dredged from the ELSC⁶ indicate a strongly depleted character relative to MORB and decreased arc geochemical signature relative to the VFR. The ELSC ends in a 65-km-wide non-transform offset from the CLSC that encompasses extinct ELSC spreading segments to the north that we infer were also formed by magma deficient spreading before they failed as the CLSC propagated southward. A refraction profile at 18° 33' S reveals 5.5-km-thick crust at these extinct segments¹¹. Zone 3 crust flanks the northern ELSC (Fig. 2) and is characterized by high adjusted mantle Bouguer anomalies, sea floor more than 1 km deeper, and crust 2–3.5 km thinner, than in zone 2. The sea-floor morphology is dominated by low-relief linear abyssal hills, in marked contrast to the more volcanic morphology of zone 2. Although seismically

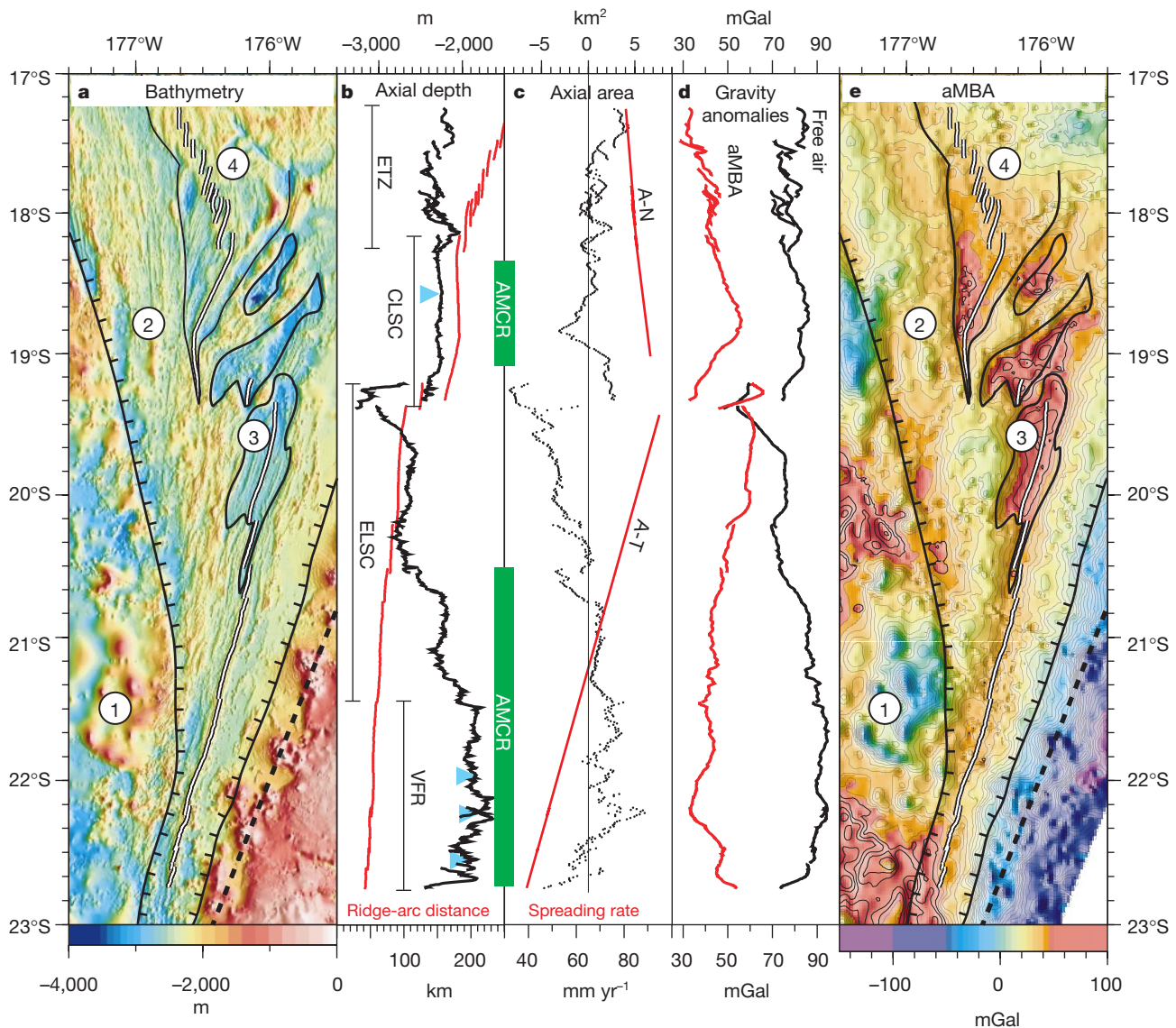


Figure 2 Map and axial profiles of the Lau spreading centres. **a**, Map view of the bathymetry. **b**, Along-axis depth profile (black line) and distance of the spreading centres from the volcanic front (red line). Green areas labelled AMCR indicate where axial magma chamber reflectors have been observed⁹. Blue triangles show active hydrothermal sites^{19,21}. **c**, Across-axis area (dots) and spreading rate variation⁵ (red line). A–T, Australia–Tonga spreading rates. A–N, Australia–Niuafo’ou spreading rates. **d**, Adjusted mantle Bouguer anomaly (aMBA; red line) and free-air anomaly (black line). The adjusted mantle Bouguer anomaly is the gravity field remaining after the effects of

bathymetry, of a 6-km-thick oceanic crustal model, and the long-wavelength gravity gradient due to the slab (see Supplementary Information) are removed from the free-air gravity anomaly. **e**, Map view of the adjusted mantle Bouguer anomaly. Warm colours indicate relatively thin crust, and cool colours relatively thick crust. Panels **a** and **e** have a tectonic interpretation showing location of the boundary (ticked line) between the older crust of the basin (1) and various outlined crustal domains (2–4) associated with the Lau spreading centres (white lines). Dashed line indicates arc volcanic front.

determined crustal thicknesses are only available for the extinct spreading segments of the ELSC, the morphology, depth, gravity, geochemistry and absence of an axial magma chamber reflector all consistently indicate that the ELSC has a decreased magma supply relative to the VFR and CLSC despite higher spreading rates.

The CLSC is located 160–185 km from the volcanic front, and 65 km west of the northern ELSC. As a result of a three-plate configuration in the northern Lau basin⁵ (Fig. 1), the CLSC has slightly slower spreading rates (90–85 mm yr⁻¹) than the northern ELSC. Nevertheless, the CLSC axis abruptly shallows and forms an axial high (~2,300 m) with low axial depth variations (Fig. 2). Lavas from the CLSC have MORB-like geochemistry without the depleted character and arc geochemical signal evident in the ELSC⁶. North of its propagating tip, the CLSC has positive inflation, adjusted mantle

Bouguer anomalies are ~20 mGal lower than the northern ELSC (Fig. 2), a bright, shallow, axial magma chamber reflector is observed from 18° 21' S–19° 05' S (ref. 9), and active hydrothermal vents²¹ occur at 18° 36' S. Relative to the CLSC, the ELSC lavas have lower or equal Fe contents (for example, $9 \pm 1\%$ FeO at 8% MgO) and have higher porosity⁶. This indicates that the shallower depth and lower adjusted mantle Bouguer anomalies on the CLSC are not due to crustal compositional (density) differences, but rather to thinner crust on the ELSC. Zone 4 crust formed as the CLSC propagated south, producing variable depths and high gravity anomalies along its pseudofaults. Depths, flank subsidence with age, and gravity anomalies away from the pseudofaults are consistent with normal (6–7 km) oceanic crustal thicknesses. Seismic refraction data confirm that crust formed at the CLSC at 18° 30' S is 7 km thick¹¹. Thus the axis and flanking crust of the CLSC has geophysical, morphologic, seismic, geochemical and hydrothermal characteristics similar to fast-spreading mid-ocean ridges.

The above observations show that the present-day axes of the Lau basin spreading centres (and the sea floor generated through time on these axes) display characteristics that vary systematically with their position over the mantle wedge—but not with spreading rate. We propose that these variations result from the migration with time of the melt source regions of the back-arc spreading centre through a mantle wedge that is variably melt enhanced/depleted spatially (Fig. 3). A general feature of intra-oceanic arc systems is a gradient in depletion of high-field-strength elements relative to MORB that increases from the back-arc toward the arc volcanic front^{22–24}. It has been shown that such a pattern exists even without back-arc spreading^{22,23} so that the mantle wedge is self-depleting, probably because water released from the slab lowers the melt solidus²⁵. This melt buoyantly rises, underplating, intruding or erupting through the upper plate, or is driven to the volcanic front by subduction-induced flow²⁶, and leaves a residual mantle depleted of a melt fraction. Corner flow induced by subduction drives the depleted layer toward the wedge corner, where increasing water concentrations from the slab promote additional melting²⁵ and maintain low viscosity, permitting the otherwise stronger^{27,28} depleted layer to continue to flow. At the mantle wedge corner, viscous coupling with the subducting slab overturns the depleted layer and carries it back beneath the back-arc basin (Fig. 3).

When spreading propagates into this depleted and hydrated wedge corner, as at the VFR, its melting regime advects the mantle upward as a consequence of plate separation and causes additional and enhanced melting from this source. It also directly draws to the spreading centre some of the melts that would otherwise contribute to arc volcanic front magmatism³. With increasing separation of the spreading centres from the volcanic front, as at the ELSC, the magmatic regimes advect less of the arc volcanic front melt and are increasingly restricted to generating melt solely by advecting the depleted wedge corner. Consequently, as this depleted corner zone is now even more depleted, having had a melt component removed by previous spreading and by volcanic front magmatism, the total amount of melt that can be produced decreases. This results in thinner and deeper crust, low inflation, absence of magma chamber reflectors, high adjusted mantle Bouguer anomalies, and the high geochemical depletion characteristics⁶ of the ELSC. Sufficiently far from the wedge corner, however, undepleted mantle must be advecting into the mantle wedge to balance the overall eastward roll-back of the Pacific slab. Such a mantle flow has been inferred on the basis of the recent change in isotopic signature of basin lavas from Pacific to Indian type²⁹. The melt source region of the CLSC intersects this undepleted mantle so that the spreading centre has typical mid-ocean-ridge characteristics for its spreading rate. Indeed, we speculate that it was the increasingly more refractory mantle beneath the ELSC (and its increasing viscosity as it separates

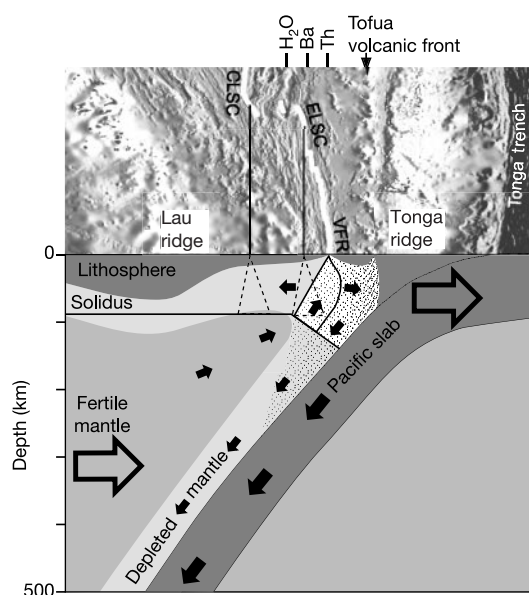


Figure 3 Model of mantle wedge control on back-arc crustal accretion. Upper panel shows an oblique view of the Lau basin bathymetry and spreading centres aligned vertically along the volcanic front, and locates the limits in arc geochemical influence determined from concentrations of H₂O, Ba and Th (ref. 6). The lower panel schematically shows a vertical cut through the mantle wedge at approximately 1:1 scale. Large open arrows indicate the roll-back of the Tonga trench and Pacific slab, and the compensating flow of mantle beneath the Lau basin. Large filled arrows show the subduction component of the Pacific slab, and small solid arrows show flow in the mantle wedge induced by the slab subduction and back-arc spreading. Stippled gradient indicates region of hydrated mantle, with water concentration increasing eastward toward the slab. Within this region, the solidus is shown depressed owing to the effect of water. Region of partial melt in the mantle is shown as the white background beneath stippling. The melting regime of the Valu Fa ridge is shown as the outlined region. It is asymmetric and distorted because of the mantle flow field³ in the wedge corner. The robust magmatic characteristics of the VFR result from the greater melt production within the hydrated mantle near the corner, and from advection of some arc melts that would otherwise supply the volcanic front. Melt extraction by the spreading centre and by arc volcanism results in a depleted mantle region (light grey). Owing to continuous slab fluid addition, the depleted layer remains weak and can be entrained in the corner flow, become overturned, and re-introduced beneath the back-arc basin. The projected positions of the ELSC and CLSC melting regimes are shown as dashed triangles. The ELSC melting regime is too far from the volcanic front to directly draw significant arc melt. It overlies primarily highly depleted mantle that is being carried into the back-arc by the slab. As a result it produces low amounts of melt, thin crust and deep sea-floor despite fast spreading rates. The projected position of the CLSC melting regime (left dashed triangle) overlies fertile mantle largely removed from slab effects. Consequently the CLSC has crustal thickness, morphology and geochemistry like those of a mid-ocean ridge.

from the slab hydration source that permits refractory mantle to flow) that may have promoted the propagation of the CLSC to replace this system.

Spreading in the Lau back-arc basin is at present well-organized, as at mid-ocean ridges^{5,12}, but, as we have shown, the supply of magma is modulated by mantle wedge compositional controls and arc melt additions. Similar controls are indicated in the Mariana trough, even though the slab there is steeper, and rates of spreading and subduction are slower³⁰. In the northern Mariana trough¹⁶, the spreading centre varies from magmatically robust, to magma starved, to normal as it separates from the volcanic front, and in the southern trough the spreading centre changes from an axial valley to an axial high as it approaches the volcanic front³⁰. These variations in relative magma supply with proximity to the volcanic front are similar to those we report here for the Lau basin, and support the generality of our model of mantle wedge control on back-arc crustal accretion.

Received 22 October 2001; accepted 14 February 2002.

- Langmuir, C. H., Klein, E. M. & Plank, T. In *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) (American Geophysical Union, Washington DC, 1992).
- Davies, J. H. & Stevenson, D. J. Physical model of source region of subduction zone volcanics. *J. Geophys. Res.* **97**, 2037–2070 (1992).
- Ribe, N. M. Mantle flow induced by back arc spreading. *Geophys. J. Int.* **98**, 85–91 (1989).
- Chiu, J.-M., Isacks, B. L. & Cardwell, R. K. 3-D configuration of subducted lithosphere in the western Pacific. *Geophys. J. Int.* **106**, 99–111 (1991).
- Zellmer, K. E. & Taylor, B. A three-plate kinematic model for Lau Basin opening. *Geochem. Geophys. Geosyst.* [online] **2**, 2000GC000106 (2001).
- Pearce, J. A. *et al.* in *Volcanism Associated with Extension at Consuming Plate Margins* (ed. Smellie, J. L.) 53–75 (Geological Society, London, 1995).
- Vallier, T. L. *et al.* Subalkaline andesite from Valu Fa Ridge, a back-arc spreading center in southern Lau Basin: petrogenesis, comparative chemistry, and tectonic implications. *Chem. Geol.* **91**, 227–256 (1991).
- Jenner, G. A., Cawood, P. A., Rautenschlein, M. & White, W. M. Composition of back-arc basin volcanics, Valu Fa ridge, Lau Basin: Evidence for a slab-derived component in their mantle source. *J. Volcanol. Geotherm. Res.* **32**, 209–222 (1987).
- Harding, A. J., Kent, G. M. & Collins, J. A. Initial results from a multichannel seismic survey of the Lau back-arc basin. *Eos* **81**, F1115 (2000).
- Turner, I. M., Peirce, C. & Sinha, M. C. Seismic imaging of the axial region of the Valu Fa Ridge, Lau Basin—the accretionary processes of an intermediate back-arc spreading ridge. *Geophys. J. Int.* **138**, 495–519 (1999).
- Crawford, W. C., Hildebrand, J. A., Dorman, L. M., Webb, S. C. & Wiens, D. A. Tonga Ridge and Lau Basin crustal structure from seismic refraction data. *J. Geophys. Res.* (in the press).
- Taylor, B., Zellmer, K., Martinez, F. & Goodliffe, A. Sea-floor spreading in the Lau back-arc basin. *Earth Planet. Sci. Lett.* **144**, 35–40 (1996).
- Scheirer, D. & Macdonald, K. C. Variation in cross-sectional area of the axial ridge along the East Pacific Rise—Evidence for the magmatic budget of a fast spreading center. *J. Geophys. Res.* **98**, 7871–7885 (1993).
- Sinha, M. C. Segmentation and rift propagation at the Valu Fa ridge, Lau Basin: Evidence from gravity data. *J. Geophys. Res.* **100**, 15025–15043 (1995).
- Hawkins, J. W. *Backarc Basins: Tectonics and Magmatism* (ed. Taylor, B.) 63–138 (Plenum, New York, 1995).
- Martinez, F., Fryer, P., Baker, N. A. & Yamazaki, T. Evolution of backarc rifting: Mariana Trough, 20°–24°N. *J. Geophys. Res.* **100**, 3807–3827 (1995).
- Wright, I. C., Parson, L. M. & Gamble, J. A. Evolution and interaction of migrating cross arc volcanism and back-arc rifting: An example from the southern Havre Trough (35°20′–37°S). *J. Geophys. Res.* **101**, 22071–22086 (1996).
- Wiedicke, M. & Collier, J. Morphology of the Valu Fa Spreading Ridge in the Southern Lau Basin. *J. Geophys. Res.* **98**, 11769–11782 (1993).
- Fouquet, Y. *et al.* Hydrothermal activity and metallogenesis in the Lau back-arc basin. *Nature* **349**, 778–781 (1991).
- Kappel, E. S. & Ryan, W. B. F. Volcanic episodicity and a non-steady state rift valley along northeast Pacific spreading centers: evidence from SeaMARC I. *J. Geophys. Res.* **91**, 13925–13940 (1986).
- Bortnikov, N. S., Fedorov, D. T. & Murav'ev, K. G. Mineral composition and conditions of the formation of sulfide edifices in the Lau Basin (southwestern sector of the Pacific Ocean). *Geol. Ore Deposits* **35**, 476–488 (1993).
- Hochstaedter, A. G. *et al.* Across-arc geochemical trends in the Izu-Bonin arc: Constraints on source composition and mantle melting. *J. Geophys. Res.* **105**, 495–512 (2000).
- Hochstaedter, A. G., Kepezhinskas, P., Defant, M., Drummond, M. & Koloskov, A. Insights into the volcanic arc mantle wedge from magnesian lavas from the Kamchatka arc. *J. Geophys. Res.* **101**, 697–712 (1996).
- Woodhead, J., Eggins, S. & Gamble, J. High field strength and transition element systematics in island arc and backarc basin basalts: Evidence for multiphase melt extraction and a depleted mantle wedge. *Earth Planet. Sci. Lett.* **114**, 491–504 (1993).
- Stolper, E. & Newman, S. The role of water in the petrogenesis of Mariana trough magmas. *Earth Planet. Sci. Lett.* **121**, 293–325 (1994).
- Spiegelman, M. & McKenzie, D. Simple 2-D models for melt extraction at mid-ocean ridges and island arcs. *Earth Planet. Sci. Lett.* **83**, 137–152 (1987).

- Phipps Morgan, J. The generation of a compositional lithosphere by mid-ocean ridge melting and its effect on subsequent off-axis hotspot upwelling and melting. *Earth Planet. Sci. Lett.* **146**, 213–232 (1997).
- Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
- Hergt, J. M. & Hawkesworth, C. J. Pb-, Sr-, and Nd-isotopic evolution of the Lau Basin: Implications for mantle dynamics during backarc opening. *Proc. ODP Sci. Res.* **135**, 505–517 (1994).
- Martinez, F., Fryer, P. & Becker, N. Geophysical characteristics of the Southern Mariana Trough, 11°50′ N–13°40′ N. *J. Geophys. Res.* **105**, 16591–16608 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank A. Harding, G. Kent and J. Collins for providing gravity and multibeam data from RV *Maurice Ewing* cruise EW9914, and A. Goodliffe and K. Zellmer for providing bathymetric data compilations and for discussions. We also thank D. Scheirer for comments on the manuscript that improved this contribution. This work was supported by the US NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.M. (e-mail: martinez@soest.hawaii.edu).

Determinants of extinction in the fossil record

Shanan E. Peters & Michael Foote

Department of the Geophysical Sciences, University of Chicago, 5734 S. Ellis Avenue, Chicago, Illinois 60637, USA

The causes of mass extinctions and the nature of biological selectivity at extinction events are central questions in palaeobiology. It has long been recognized, however, that the amount of sedimentary rock available for sampling may bias perceptions of biodiversity^{1–7} and estimates of taxonomic rates of evolution^{5–8}. This problem has been particularly noted with respect to the principal mass extinctions^{5–12}. Here we use a new compilation of the amount of exposed marine sedimentary rock to predict how the observed fossil record of extinction would appear if the time series of true extinction rates were in fact smooth. Many features of the highly variable record of apparent extinction rates within marine animals can be predicted on the basis of temporal variation in the amount of exposed rock. Although this result is consistent with the possibility that a common geological cause determines both true extinction rates and the amount of exposed rock, it also supports the hypothesis that much of the observed short-term volatility in extinction rates is an artefact of variability in the stratigraphic record.

The highly variable rock record may distort the apparent timings of taxonomic origination and extinction both locally, at the outcrop scale, and globally, in taxonomic databases. For example, at outcrop and basin scales, principles of sequence stratigraphy imply that last occurrences of taxa should cluster artificially at temporal gaps and shifts in depositional environments that result from basin in-filling, eustatic sea-level variation, and crustal uplift and subsidence^{7,12–14}. To evaluate the relationship between global preservation potential, as reflected by rock availability, and the last sampled occurrences of taxa, we first compared the global, stage-level records of apparent genus extinction with our estimate of exposed marine rock (See Methods). Genus data are from Sepkoski's unpublished compen-

dium, and our measure of exposed rock is the number of marine sedimentary formations. The number of named formations is a useful proxy for the amount of rock because it reflects several relevant aspects of sampling, including exposed outcrop area (see Methods).

Figure 1 shows the stage-level records of genus extinction and the number of marine formations. The two time series are positively correlated in absolute value (Fig. 1a, $r = 0.304$) and when they are de-trended by analysing stage-to-stage changes (that is, first differences) in each (Fig. 1b, $r = 0.238$, $P = 0.056$). These correlations are weak, but there are important reasons to expect an imperfect relationship between the amount of rock and apparent extinction in the same time interval. Notably, many taxa that actually became extinct during an interval of time with a poor record are unlikely to be preserved during that interval and will therefore have spurious last occurrences, producing an artificially high extinction rate, in the

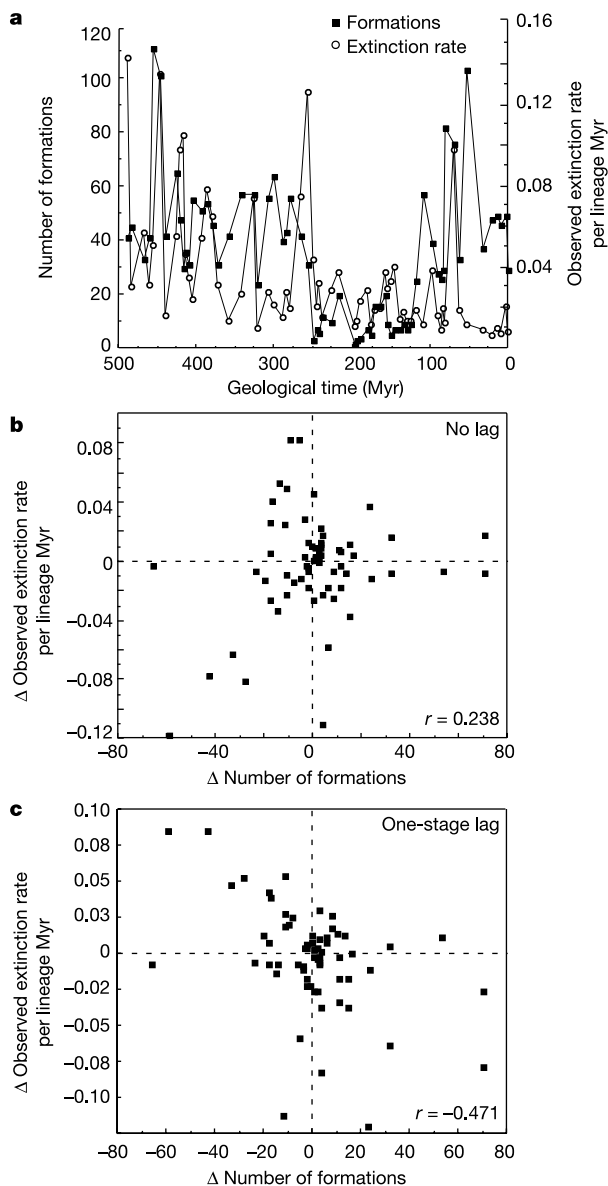


Figure 1 Observed extinction rate per lineage Myr (see Methods) and the number of marine formations in the combined data set. **a**, Time series. **b**, First differences (value for a stage minus the value for the previous stage) from **a**. **c**, Change in observed extinction rate from stage $i - 1$ to stage i versus change in amount of rock from stage i to stage $i + 1$. Myr, million years.

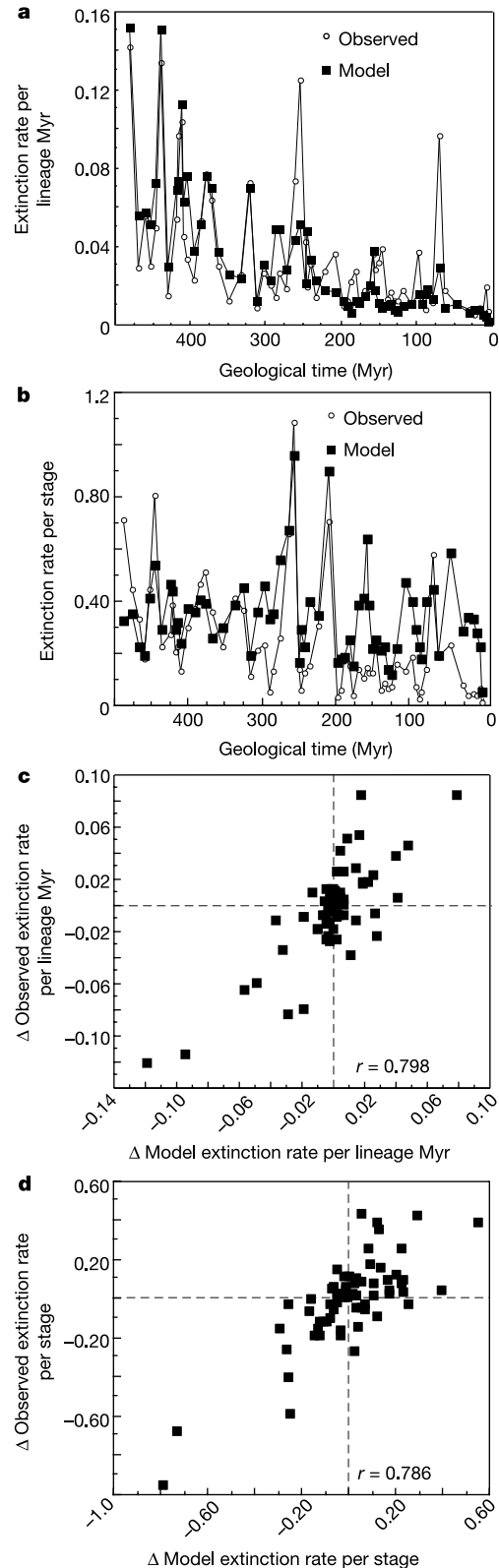


Figure 2 Observed extinction rates and predicted apparent extinction rates based on filtering a smooth extinction history through the empirically calibrated time series of preservation rates using the combined compilation. **a**, Time series for model with variable stage lengths and declining extinction rates. **b**, Time series for model with constant stage lengths and constant extinction rates. **c**, First differences from **a**. **d**, First differences from **b**. Correlations are significant ($P < 0.0005$) according to a formation randomization test. Observed extinction rates in **a** and **b** are identical, except that **a** expresses per capita rates per lineage Myr, whereas **b** expresses per capita rates per stage.

Table 1 Linear product-moment correlation coefficients

Formation compilation	Extinction model	Stage lengths	Recovery rate per formation (<i>R</i>)			
			<i>R</i> = 0.005	<i>R</i> = 0.01	<i>R</i> = 0.02	<i>R</i> = 0.04
Combined	Declining	Based on timescale	0.798	0.798	0.779	0.739
		Assumed constant	0.738	0.778	0.811	0.821
	Constant	Based on timescale	0.768	0.765	0.743	0.696
		Assumed constant	0.780	0.786	0.790	0.775
Lexicon*	Declining	Based on timescale	0.716	0.725	0.724	0.703
		Assumed constant	0.583	0.629	0.681	0.731
	Constant	Based on timescale	0.641	0.649	0.652	0.634
		Assumed constant	0.665	0.673	0.691	0.709
Literature†	Declining	Based on timescale	0.832	0.843	0.840	0.814
		Assumed constant	0.714	0.751	0.773	0.773
	Constant	Based on timescale	0.811	0.809	0.796	0.757
		Assumed constant	0.812	0.807	0.797	0.779

Data are product-moment correlation coefficients for de-trended empirical extinction rate time series and the predicted time series of extinction rates for different formation compilations, and for various model calibrations. The declining extinction model predicts true extinction rates that are equal to exponential fits to the empirical data. The constant extinction model assumes that rates are constant at the Phanerozoic mean. Stage lengths are assumed to be equal up-to-date calibrations^{27,28} or are assumed to be constant at unit duration. The probability of recovery per formation (*R*) was also allowed to vary. All correlations are significant at $P < 0.0005$ according to a formation randomization test.

* Taken from ref. 25.

† Taken from data published in the *Journal of Paleontology* and *Palaeontology* (see text for details).

preceding interval. Similarly, intervals with exceptionally good rock records will be preceded by artificially low extinction rates¹⁵.

To test the importance of this offset effect, we computed the correlation between the change in amount of rock from stage *i* to stage *i* + 1 versus the change in observed extinction rate from stage *i* - 1 to stage *i* (that is, first differences at a lag of one stage; Fig. 1c). As expected, this correlation is negative ($r = -0.471$, $P < 0.0001$) and stronger than that between the amount of rock and apparent extinction in the same stage. What is important to note in Fig. 1c is that substantial increases in observed extinction rate are generally followed by declines in exposed sedimentary rock, and major decreases in observed extinction rate tend to be followed by increases in exposed rock.

Although a negative correlation between extinction and preservation at a lag of one time interval constitutes *prima facie* evidence for artificial volatility in the extinction record, calculating correlations between the amount of rock and extinction rates does not

fully capture the potential effects of variability in the stratigraphic record. This is because the entire time series of the quality of preservation, as well as true variation in origination and extinction rates, affects the apparent extinction rate for each stage¹⁵. To assess the extent to which variation in extinction rates can be predicted by variation in the amount of rock, we therefore used a mathematical model to determine the expected pattern of extinction resulting from filtering a smooth extinction time series through the empirically calibrated time series of preservation. This model posits that true rates of extinction have varied smoothly over the Phanerozoic eon, and that observed volatility in extinction is partly an artefact of temporally heterogeneous preservation.

Figure 2 compares the empirical history of observed extinction rates with the model-predicted histories for two endmember models: one that assumes variable stage lengths and declining extinction rates (Fig. 2a, c), and one that assumes constant stage lengths and constant extinction rates (Fig. 2b, d). Within each endmember calibration, many features in the observed extinction data and the models agree. Although details of the model predictions vary, model–data correlations are similar for a wide range of model calibrations and for each compilation of formations (Table 1).

To test the statistical significance of the de-trended model–data correlations shown in Fig. 2 and Table 1, we generated null distributions of correlation coefficients by randomizing the formation time series and then recalculating the correlation between model and observations. Null distributions for the models shown in Fig. 2 are presented in Fig. 3. The null distribution for the variable stage-length scheme is not zero-centred. The offset reflects the shared timescale used to calculate both model and observed extinction rates. By contrast, the null distribution for the constant stage-length model is zero-centred. In every case, the observed formation data yield a correlation stronger than the distribution of 2,000 correlations generated by randomizing the formation data. Estimating the proportion of variance in apparent extinction rates that is potentially explained by the model would require knowledge of which model and which parameter values for that model are most realistic. Although these requirements cannot yet be met, we have shown results for a broad range of models and calibrations likely to represent reality. Regardless of the model assumed and the parameter values used to calibrate the model, the observed time series of formation numbers has a statistically significant ($P < 0.0005$) ability to predict variation in apparent extinction rate.

The extent to which deviations between our model and data reflect true biological anomalies, errors in Sepkoski's taxonomic data, inadequacies in our measure of rock quantity, and other

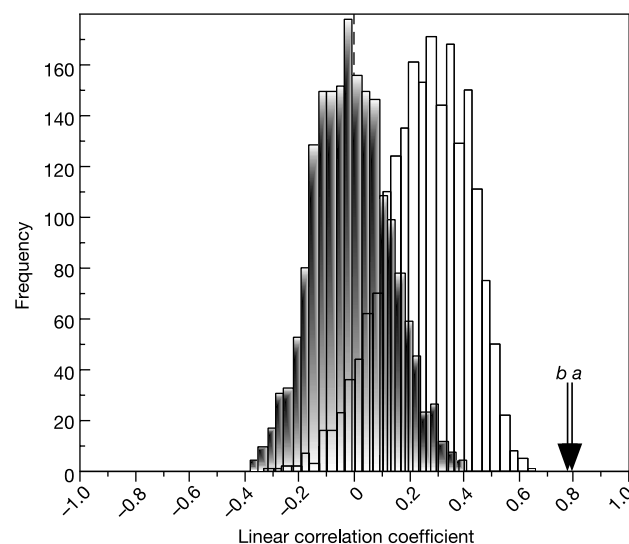


Figure 3 Null distributions of linear product-moment correlation coefficients generated by randomizing the combined formation time series. Open distribution shows model with variable stage lengths and declining extinction rates (a); shaded distribution shows model with constant stage lengths and constant extinction rates (b). Arrows indicate observed correlation coefficients, which are completely outside the range of their respective null distributions ($P < 0.0005$).

aspects of model calibration, is as yet unknown. Model results should therefore be interpreted as statistical descriptions of the potential effects of variable preservation on short-term volatility in apparent extinction rates, rather than as predictions of the absolute values of extinction rates. For example, whether the end-Permian (approximately 251 million years (Myr) ago) and end-Triassic (about 200 Myr ago) extinction peaks deviate from model predictions depends on whether true extinction rate is modelled as constant or declining (Fig. 2). Similarly, the scaling of the model–data relationship depends on the hypothesized time series of true extinction and the assumed per-formation probability of preservation. For the results shown in Fig. 2a, for example, the regression of de-trended observed rates on de-trended model rates has a slope of 1.04. Increasing or decreasing the assumed probability of preservation (see Methods) leads to a corresponding change in this slope. For the 48 model calibrations in Table 1, slopes range from 2.24 to 0.63, with a mean of 1.05 and a median of 0.91. Because the amplitude of variation in rates varies with model parameters, we choose to focus on the strength of the model–data correlation, which measures the agreement in the temporal pattern of rates. In this respect, our results are insensitive to model calibrations and to the hypothesized time series of extinction.

There are two explanations for the agreement between observed and model-predicted de-trended rates of genus extinction. The first is that the apparent volatility in extinction is largely spurious. Although our results suggest that there is an artificial overprint on the empirical data, some extinction events, such as the end-Permian, may involve a greater loss of genera than expected on the basis of known gaps in the record¹⁶. Despite the raw magnitude of such extinction events, it is possible that spurious termination of lineages by taxonomic practice, or pseudo-extinction, is prevalent near gaps in the stratigraphic record and at times of apparent increase in extinction rate¹⁷. Although the magnitude of pseudo-extinction at apparent mass extinctions is generally unknown, this effect would only serve to strengthen the artefact hypothesis.

The second explanation for the match between model and data is that a common cause affects both extinction and the amount of rock preserved and exposed at the Earth's surface. The most obvious candidate is sea level, a decline in which is expected to cause a drop in the number of preserved formations and potentially an increase in true extinction through the elimination of shallow marine habitats and overall reduction in habitable area^{18,19}. Although sea-level change may be a viable mechanism for extinction, the relationship between sea level, extinction, and preservation is potentially more complex than suggested by a simple common-cause hypothesis. For example, sea-level increase has been invoked as a causal agent of extinction through the spread of anoxic environments¹⁹, and increases in sea level can also have deleterious impacts on stratigraphic sampling⁷.

Regardless of the reason for the correlation, the agreement between observed rates of extinction and those predicted from stratigraphic data must affect our understanding of large-scale evolution in the Phanerozoic. If the correlation arises because the record has imposed a strong bias on fossil data, then the marine biosphere may have been substantially more stable than previously supposed. Testing this hypothesis and correcting biases in the record will be facilitated by additional work on sampling standardization²⁰, model-based fitting of rates²¹, phylogenetic standardization of taxonomy across stratigraphic boundaries¹⁷, and more complete, global compilations of the amount of sedimentary rock, its quality, and its stratigraphic context. If, instead, extinction rates and the amount of preserved sedimentary rock share a common cause, then our results may serve to constrain the mechanisms and timing of extinction, reducing the likelihood, for example, that bolide impacts^{22–24} are the principal cause of most extinction events.

Although we suspect that some extinction episodes may ultimately prove too severe to be explained entirely as stratigraphic artefacts, our analyses provide a measure of quantitative support for the hypothesis^{5,7} that a significant component of the observed variation in rates of genus extinction may be spurious. □

Methods

Marine formation data

We used ref. 25 to compile the number of named lithostratigraphic rock units in the United States and its territories for each of 77 conventional Phanerozoic stratigraphic intervals, principally stages, having an average duration of 7.1 Myr. Formation was chosen as the unit of sedimentary rock because it has a standardized definition (lithologically distinct and mappable at a scale of 1:24,000) and is not erected on the basis of fossil content. A total of 1,051 marine formations could be assigned to stratigraphic stages.

To supplement our compilation of formations from ref. 25 with world-wide data, we included a formation compilation derived by systematically surveying the palaeontological literature. This continuing survey currently covers data published over the past 10 years in the *Journal of Paleontology* and the past three years in *Palaeontology*. Each marine formation was counted only once. When de-trending data by analysing stage-to-stage changes (that is, first differences), we find that the literature- and ref. 25-derived time series are positively correlated ($r = 0.751$, $P < 0.0001$). A total of 56.4% of 830 formations in the literature compilation and 27.5% of 1,700 formations in the combined compilation are from outside the United States. The de-trended time series of these formations outside of the United States is positively correlated with that for the United States-only formations from ref. 25 ($r = 0.456$, $P < 0.001$). Thus, our formation tabulations repeat similar patterns when compiled from different sources and for different geographical regions.

In addition to testing for agreement between formation compilations, we tested our data against independent estimates of global marine outcrop area. The Paleogeographic Atlas Project provided outcrop area data for 12 Mesozoic and Cenozoic stages. Area was measured as the number of equal-area grids occupied by at least one marine outcrop—grid size was 0.5×0.5 degrees at the equator, but other grid sizes yield similar results. The three formation compilations are positively correlated with global outcrop area (Spearman coefficients range from 0.752 to 0.813 and are significant at $P < 0.02$). In addition to providing an estimate of absolute outcrop area, formations may also reflect other relevant aspects of sampling including research effort and range of habitat preservation⁶.

Extinction data

We tabulated empirical rates of genus extinction using data on animals and animal-like protists from the global compendium of Sepkoski²⁶. In total, we used 30,379 marine genera with both first and last fossil appearances resolved at least to the stage level. Extinction was measured as the instantaneous, per capita rate of extinction per lineage per million years (lineage Myr)¹⁵. Stage lengths were based on a number of up-to-date sources^{27,28}. Because Cambrian formations could rarely be confidently assigned to stages using ref. 25, we present results only for the post-Cambrian Phanerozoic.

Extinction model

For the hypothesized declining extinction model, we fitted an exponential decline through the observed rates for the Palaeozoic era and again through the Mesozoic and Cenozoic eras²⁹. We chose to fit smooth curves to the observed data because the long-term empirical rate is expected to be approximately equal to the true long-term rate, no matter how biased short-term fluctuations may be by incomplete sampling¹⁵. This fitting procedure is an imperfect approximation, and the observed decline has even been claimed to reflect bias in taxonomic procedure³⁰. We nevertheless find consistent results if we ignore the long-term trend in extinction rates and assume that true extinction is constant at the Phanerozoic mean (Fig. 2 and Table 1). For simplicity, we assumed that origination and extinction rates are equivalent for each stage, but we obtain consistent results if we assume instead that origination rates are equal to their observed values.

We next used our formation data to derive a time series of instantaneous preservation rates per genus per Myr—by preservation we mean all events leading to the inclusion of a genus in a taxonomic database, including burial in sediments, collection and publication. We assumed a uniform probability of recovery per genus per formation. This is a simplification, but, provided that recovery probabilities vary stochastically over time, with or without a long-term trend¹⁵, this simplification should not bias our results. If R is the probability of recovery per formation, x is the rate of preservation per genus per Myr, F is the number of formations, and Δt is the duration of the stage, then $x = -F \ln(1 - R)/\Delta t$. This is because the probability that a genus ranging through a time interval will not be sampled is equal to $\exp(-x\Delta t)$ and to $(1 - R)^F$. Spreading a given number of formations over a longer interval of time therefore yields a lower rate of fossil recovery per unit time.

The value of R was chosen to yield a long-term average value of x comparable to those that have been empirically determined previously for the same genus data, namely about 0.05 per genus per Myr (ref. 21). However, correlation coefficients are consistent for a wide range of values of R (Table 1).

The goal of the forward model is to predict apparent rates of extinction given the temporal pattern of preservation and of hypothesized rates of origination and extinction. Let $p(T)$, $q(T)$ and $x(T)$ be time-specific rates of origination, extinction and preservation per genus per Myr, respectively. Consider an interval of time from $T = t$ to $T = t + \Delta t$.

Let X_b be the number of genera observed to cross the beginning boundary of the time interval, and X_{bt} be the number observed to cross both the beginning and terminal interval boundaries. Let N_b and N_{bt} be the corresponding true numbers. N_b is proportional to $\exp\{\int_0^t p(T) - q(T) dT\}$. If $\bar{q}$ is the true mean extinction rate through the interval, then N_{bt} is equal to $N_b \exp(-\bar{q} \Delta t)$; thus $N_{bt}/N_b = \exp(-\bar{q} \Delta t)$ and $\bar{q} = -\ln(N_{bt}/N_b)/\Delta t$. The corresponding apparent rate of extinction is equal to $-\ln(X_{bt}/X_b)/\Delta t$. X_b and X_{bt} can be predicted from N_b and N_{bt} and the probabilities of being preserved before and after a given boundary, which in turn depend on the time series of $p(T)$, $q(T)$ and $x(T)$. A more detailed explanation of the model is given in ref. 15, and model source code is provided in the Supplementary Information.

Received 10 September 2001; accepted 17 January 2002.

1. Raup, D. M. Taxonomic diversity during the Phanerozoic. *Science* **177**, 1065–1071 (1972).
2. Raup, D. M. Species diversity in the Phanerozoic: an interpretation. *Paleobiology* **2**, 289–297 (1976).
3. Miller, A. I. & Foote, M. Calibrating the Ordovician radiation of marine life: implications for Phanerozoic diversity trends. *Paleobiology* **22**, 304–309 (1996).
4. Miller, A. I. in *Deep Time: Paleobiology's Perspective* (eds Erwin, D. H. & Wing, S. L.) *Paleobiology* **26**(4), (Suppl.), 53–73 (2000).
5. Smith, A. B. Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies. *Phil. Trans. R. Soc. Lond. B* **356**, 351–367 (2001).
6. Peters, S. E. & Foote, M. Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology* **27**, 583–601 (2001).
7. Smith, A. B., Gale, A. S. & Monks, N. E. A. Sea-level change and rock-record bias in the Cretaceous: a problem for extinction and biodiversity studies. *Paleobiology* **27**, 241–253 (2001).
8. Patterson, C. & Smith, A. B. Is the periodicity of extinction a taxonomic artefact? *Nature* **330**, 248–251 (1987).
9. Raup, D. M. & Sepkoski, J. Jr Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503 (1982).
10. Pease, C. M. Biases in total extinction rates of fossil taxa. *J. Theor. Biol.* **130**, 1–7 (1988).
11. MacLeod, N. & Keller, G. Hiatus distributions and mass extinctions at the Cretaceous/Tertiary boundary. *Geology* **19**, 497–501 (1991).
12. Ross, C. A. & Ross, J. R. P. Foraminiferal zonation of late Paleozoic depositional sequences. *Mar. Micropaleontol.* **26**, 469–478 (1995).
13. Holland, S. M. The stratigraphic distribution of fossils. *Paleobiology* **21**, 92–109 (1995).
14. Holland, S. M. in *Deep Time: Paleobiology's Perspective* (eds Erwin, D. H. & Wing, S. L.) *Paleobiology* **26**(4), (Suppl.) 148–163 (2000).
15. Foote, M. in *Deep Time: Paleobiology's Perspective* (eds Erwin, D. H. & Wing, S. L.) *Paleobiology* **26**(4), (Suppl.) 74–102 (2000).
16. Raup, D. M. Cohort analysis of generic survivorship. *Paleobiology* **4**, 1–15 (1978).
17. Fortey, R. A. There are extinctions and extinctions: examples from the Lower Palaeozoic. *Phil. Trans. R. Soc. Lond. B* **325**, 327–332 (1989).
18. Johnson, J. G. Extinction of perched faunas. *Geology* **2**, 479–482 (1974).
19. Hallam, A. & Wignall, P. B. Mass extinctions and sea level changes. *Earth Sci. Rev.* **48**, 217–250 (1999).
20. Alroy, J. *et al.* Effects of sampling standardization on estimates of Phanerozoic marine diversification. *Proc. Natl Acad. Sci. USA* **98**, 6261–6266 (2001).
21. Foote, M. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. *Paleobiology* **27**, 602–630 (2001).
22. Alvarez, W. *et al.* Extraterrestrial cause for the Cretaceous–Tertiary extinction: experimental results and theoretical interpretation. *Science* **208**, 1095–1108 (1980).
23. Becker, L. *et al.* Impact event at the Permian–Triassic boundary: evidence from extraterrestrial noble gases in fullerenes. *Science* **287**, 443–446 (2001).
24. Raup, D. M. Large-body impact and extinction in the Phanerozoic. *Paleobiology* **18**, 80–88 (1992).
25. Keroher, G. C. *et al.* Lexicon of geologic names of the United States for 1936–1960. *US Geol. Surv. Bull.* **1200**, 1–1434 (1967).
26. Sepkoski, J. Jr in *Global Events and Event Stratigraphy* (ed. Walliser, O. H.) 35–42 (Springer, Berlin, 1996).
27. Bowring, S. A. & Erwin, D. H. A new look at evolutionary rates in deep time: uniting paleontology and high-precision geochronology. *GSA Today* **8**, 1–6 (1998).
28. Golonka, J. & Kiessling, W. in *Phanerozoic Reef Patterns* (eds Kiessling, W., Flügel, E. & Golonka, J.) 11–20 (SEPM Special Publication 27, Tulsa, 2002).
29. Van Valen, L. A resetting of Phanerozoic community evolution. *Nature* **307**, 50–52 (1984).
30. Smith, A. B. & Patterson, C. The influence of taxonomic method on the perception of patterns of evolution. *Evol. Biol.* **23**, 127–216 (1988).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank R. H. De Simone, E. G. Hunt, S. M. Kidwell, A. McGowan and A. M. Ziegler for discussions. D. B. Rowley provided rock area data. We also thank R. H. De Simone, J. Huss, D. Jablonski, A. I. Miller, D. M. Raup and A. B. Smith for reading the manuscript. This work was supported by the US Environmental Protection Agency, the US National Science Foundation, the Paleontological Society, and the Society of Sigma Xi.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.P. (e-mail: sepeters@midway.uchicago.edu).

Patterns of colonization in a metapopulation of grey seals

Oscar E. Gaggiotti^{*†}, Felicity Jones[‡], William M. Lee^{*}, William Amos^{*}, John Harwood[§] & Richard A. Nichols^{||}

^{*} University of Cambridge, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK

[‡] Institute of Cell, Animal, and Population Biology, University of Edinburgh, Ashworth Annexe, Kings' Buildings, West Mains Road, Edinburgh EH9 3JT, UK

[§] NERC Sea Mammal Research Unit, Gatty Marine Laboratory, University of St Andrews, St Andrews, Fife KY16 8LB, UK

^{||} School of Biological Sciences, Queen Mary & Westfield College, University of London, Mile End Road, London E1 4NS, UK

[†] Present address: Metapopulation Research Group, Department of Ecology and Systematics, Division of Population Biology, PO Box 65, FIN-00014, University of Helsinki, Finland

The colonization of a new habitat is a fundamental process in metapopulation biology¹, but it is difficult to study. The emigration of colonists from established populations might be induced by resource competition owing to high local population density^{2,3}. Migration distances are also important because they determine the frequency and scale of recolonization and hence the spatial scale of the metapopulation⁴. Traditionally, these factors have been investigated with demographic approaches that are labour-intensive and are only possible in amenable species. In many cases, genetic differentiation is minimal, preventing traditional genetic approaches from identifying the source of colonists unambiguously. Here we present a bayesian approach that integrates genetic, demographic and geographic distance data. We apply the method to study the British metapopulation of grey seals, which has been growing at 6% per year over the last few decades⁵. Our method reveals differential recruitment to three newly founded colonies and implicates density-dependent dispersal in metapopulation dynamics by using genetic data.

One-third of the world population of grey seals breeds at about 50 colonies around the British Isles, mostly on offshore islands to the north and west of Scotland. Although the population is growing exponentially, individual colonies exhibit diverse dynamics, including fluctuation around a long-term trend, exponential or logistic increase, decrease to extinction and a few recent colonizations. Grey seals breed colonially in autumn. Each female's single offspring is recruited to the adult population at the age of 5–7, and females seem to show strong philopatry⁶. The Orkney Isles (Fig. 1) comprise

Table 1 Inferences for contribution of sources to founding groups, x

Source	Mean percentage contribution (confidence interval)		
	Stronsay	Copinsay	Calf of Eday
Faray	14.91 (5.44, 26.57)	22.06 (7.55, 41.68)	24.54 (9.90, 43.28)
Holm of Huip	32.57 (17.61, 48.65)	14.03 (5.30, 25.03)	12.85 (3.65, 23.57)
Holm of Spurness	25.63 (13.08, 39.27)	17.52 (7.14, 29.00)	20.64 (9.70, 34.94)
Muckle Greenholm	12.14 (1.64, 23.34)	21.29 (9.79, 35.31)	13.59 (4.67, 24.30)
Ruskhalm	9.52 (2.68, 18.15)	8.63 (1.80, 17.65)	11.59 (3.03, 20.95)
Stroma	2.57 (0.00, 13.19)	7.89 (1.42, 16.83)	7.72 (0.01, 20.21)
Swona	2.66 (0.00, 12.83)	8.58 (0.18, 18.98)	9.07 (0.03, 24.75)

Confidence intervals correspond to the 95% equal-tailed intervals of the posterior distributions for x.

50 islands lying off northeast Scotland and provide a natural setting for the study of colonization patterns in a metapopulation. Of the 21 breeding colonies in Orkney, three have gone extinct in the last 40 years and two of these have been recolonized. We have focused on three that were founded or recolonized in about 1992 (Stronsay, Copinsay and Calf of Eday) and seven well-established colonies that were the most likely sources of colonists. In particular, we wished to investigate the source of the colonists that founded these three colonies, to determine whether colonization events are triggered by the saturation of breeding sites at successful colonies. A minimum of 150 individual pups were sampled in each colony and scored for nine highly variable microsatellite loci. All samples from the recently founded colonies were collected sufficiently soon after founding to be sure that they were from the F_1 generation.

Empirical studies aimed at determining the source of colonists by using Wright's F -statistics have lacked resolution^{7–9}. Here we show that, despite the minimal genetic differentiation between populations (average pairwise standardized variance $F_{ST} = 0.0029$), it is still possible to obtain information on the contribution of different sources to the colonists and to address ecological questions on the factors that affect colonization. The bayesian/Markov chain Monte Carlo (MCMC) approach (see Methods) uses a likelihood function that generalizes the stock identification method (GSI)¹⁰ to consider F_1 individuals and to allow for assortative mating between seals that originated from the same source population. Similar approaches have been used to distinguish immigrant individuals^{11,12} and to estimate attributes of the population structure¹². In our case we extend it to a direct estimation of parameters of particular ecological interest. These are the probability that individuals from the same colony mate assortatively, w , and a vector

of genetic mixture coefficients that provides the proportionate contribution of each source to the founding groups, $\mathbf{x} = \{x_i\}$ ($\sum_i x_i = 1$). Two further ecological parameters (R , the rate of decay with distance, and S , the contribution of productivity) specify the contribution of density-dependent effects and of geographic distance to the probability distribution for $\mathbf{x}$. For each potential source we have estimated the relative production of pups that do not re-establish as adults in their natal colony, π_i . This estimate is obtained from the deceleration in the growth of the colony (see Methods). The parameter S quantifies the relative importance of π_i as a predictor of the source of colonists, and R quantifies the fall-off in contribution with distance.

The likelihood function extracts the genetic linkage disequilibrium information present in admixed populations and is constructed by first defining the probability of an individual genotype, k , given the parameter values, $P(k|\mathbf{w}, \mathbf{x})$:

$$P(k|\mathbf{w}, \mathbf{x}) = w \sum_i x_i \left(\prod_l \delta p_{il,a_{1k}} p_{il,a_{2k}} \right) + (1-w) \left[\sum_i x_i^2 \left(\prod_l \delta p_{il,a_{1k}} p_{il,a_{2k}} \right) + \sum_{i \neq j} x_i x_j \left(\prod_l p_{il,a_{1k}} p_{jl,a_{2k}} + \gamma p_{il,a_{2k}} p_{jl,a_{1k}} \right) \right] \quad (1)$$

where $p_{il,a_{nk}}$ is the frequency of the n th ($n = 1, 2$) allele at the l th

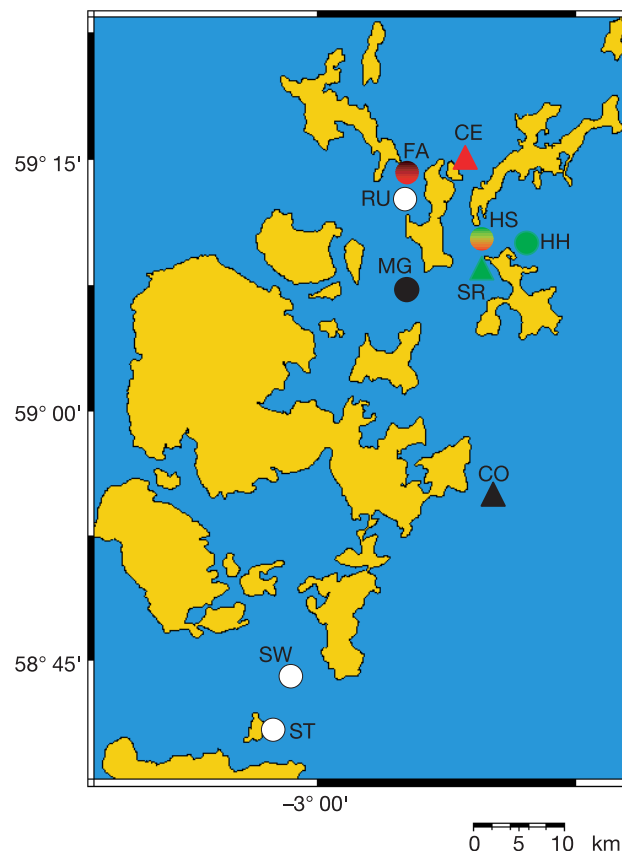


Figure 1 Location of the colonies of grey seals studied. Triangles identify new populations; circles identify potential source colonies. Each new colony and its two main providers of colonists have the same colour: red, Calf of Eday (CE); green, Stronsay (SR);

black, Copinsay (CO). Source colonies that are the main contributors to two of the new colonies have been coloured accordingly. See Methods for abbreviations of source colonies.

locus of individual k in population i and

$$\delta = \begin{cases} 1 & \text{if } a_{1lk} = a_{2lk} \\ 2 & \text{if } a_{1lk} \neq a_{2lk} \end{cases} \text{ and } \gamma = \begin{cases} 0 & \text{if } a_{1lk} = a_{2lk} \\ 1 & \text{if } a_{1lk} \neq a_{2lk} \end{cases}$$

The first term on the right-hand side of equation (1) represents the probability of genotype k given that its parents mated assortatively; the second term represents the same probability given that the parents mated at random. Using the above expression we define the likelihood of w and x given the data, y ,

$$L(w, x|y) = \Pr(y|w, x) = \prod_{k=1}^n \Pr(k|w, x) \quad (2)$$

where n is the number of individuals sampled from the newly founded colony.

We use the approach of Spiegelhalter *et al.*¹³ to compare the full model, which includes an effect of both distance and demographic factors, with simpler models that contain one or neither. This scheme makes use of bayesian measures of model complexity (p_D , the 'effective number of parameters') and fit (the posterior mean deviance). More complex models might be preferred if they give a sufficient improvement of fit, or equivalently the preferred model will have the lower value of the deviance information criterion

Table 2 DIC scores estimated from the McMC output for each model

Model	DIC score		
	Stronsay	Copinsay	Calf of Eday
Null model	6,048	6,006	6,355
S only	6,047	6,003	6,352
R only	6,044	6,003	6,354
S and R	6,044	6,002	6,352

The standard errors of the estimates were all in the range 0.040–0.017. DIC, deviance information criterion.

(DIC), which is the sum of the two measures. The DIC scores were estimated from the output of McMC chains for the appropriate model (see Methods).

In all three colonization events we found large variations in the estimated proportion of colonists from different putative sources, with the two most important sources contributing 43–59% (Table 1). Both geographic distance from sources to new colonies and demographic factors (size and dynamics of source populations) have an effect (Fig. 2). The relative importances of distance and productivity in each of the colonization events seem to have been determined by the location of potential sources around the vacant site (Fig. 1). Copinsay is more or less equidistant from all potential sources, reducing the evidence of any association with distance (Fig. 2a), and leading to similar DIC estimates for the models with and without an effect of distance (models with and without R in Table 2). Conversely, some potential source colonies are much closer than others to Stronsay and main contributors are the closest. Consequently a strong association with distance is apparent in Fig. 2b, and the relationship and the models with R are preferred (Table 2). For Calf of Eday the results indicate that both population density and distance act concurrently (Fig. 2, Table 2). In each case the preferred models include the full model with an effect of both demography and distance, which performs better than the null model. A difference in DIC of more than 3 indicates 'considerably less support' for the null model¹³.

The time series of pup production at individual colonies (see Supplementary Information) indicate that the most important sources are of medium size (Muckle Greenholm and Holm of Spurness) or large size (Faray and Holm of Huip) and currently have low or zero rates of increase. The least important contributors are either small (Ruskholm) or are still growing rapidly (Stroma and Swona). This pattern could be explained by females having a strong preference for their natal site whenever there is sufficient pupping space for them to do so.

The method developed in this paper has allowed us to combine information across loci to exploit the information in the patterns of linkage disequilibrium created by population subdivision, despite the very minor differences in allele frequency. By combining the genetic data with demographic and geographic information it proved possible to generate inferences about colonizing behaviour. Density-dependent migration has frequently been invoked in theoretical studies but is difficult to substantiate with the use of demographic data; collecting the evidence can be labour-intensive and is only possible in amenable species. Multilocus molecular genetic methods provide an easier way of challenging these models with data from natural populations. In particular, this is the first genetic study to implicate density-dependent emigration in the dynamics of a metapopulation. □

Methods

The seven potential source colonies included in this study are Faray (FA), Holm of Huip (HH), Holm of Spurness (HS), Muckle Greenholm (MG), Ruskholm (RU), Swona (SW) and Stroma (ST).

To estimate the parameters of the model described by the likelihood function and to combine genetic and demographic information we use a hierarchical bayesian approach

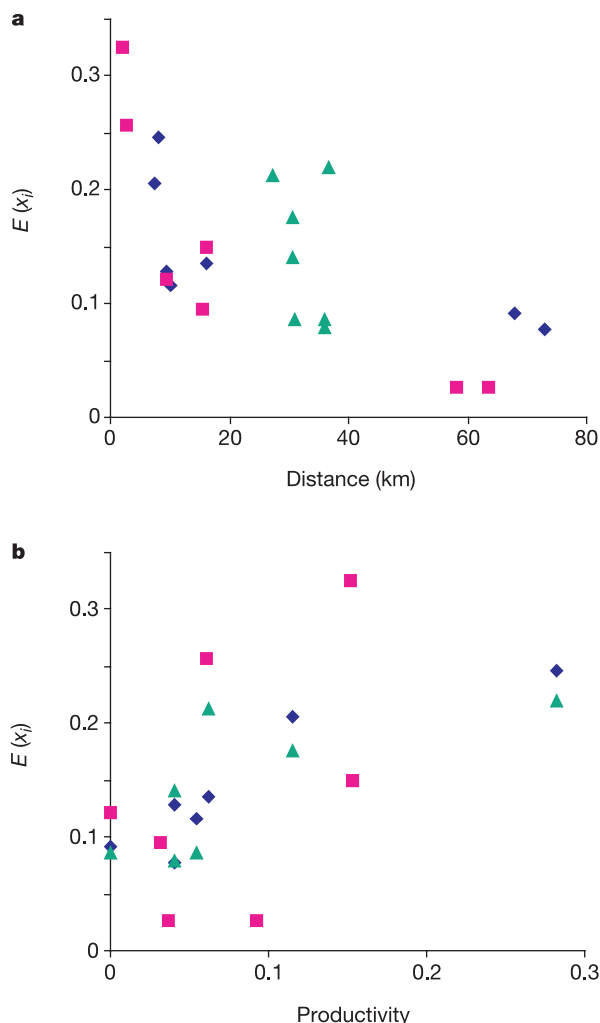


Figure 2 The estimated contribution of different source colonies to the founding groups as a function of geographic distance (**a**) and productivity (**b**). Green triangles, Copinsay; dark pink squares, Stronsay; dark blue diamonds, Calf of Eday.

that allows the expected contribution of each source population to the founding group to be a function of its productivity and its distance from the new colony,

$$E(x_i) = \frac{[(1-S) + S\pi_i]e^{-R\delta_i}}{\sum_j [(1-S) + S\pi_j]e^{-R\delta_j}}$$

where R is the rate of decay with distance δ_i , and S is the contribution of productivity π_i . Distance between sources and new colonies is measured along the path that a seal would use to move between colonies. Productivity is obtained from the time series for pup production (see Supplementary Information) and describes both how strong the density-dependent effects within a source are and how large the source population is. It is obtained as

$$\pi_i = \Phi_i(\Delta_m - \Delta_i), i = 1, 2, \dots, s$$

where

$$\Delta_i = \frac{1}{(t_{c-2} - t_{c+2} + 1)} \sum_{j=t_{c-2}}^{t_{c+2}} N_{ij}$$

is the growth rate of source i per individual during the period $[t_{c-2}, t_{c+2}]$ (t_c is the year in which the colonization event occurred and N_{ij} is the number of pups born in colony i on year j), Δ_m is the maximum growth rate per individual, and

$$\Phi_i = \frac{\sum_{j=t_{c-2}}^{t_{c+2}} N_{ij} / (t_{c-2} - t_{c+2} + 1)}{\sum_{i=1}^s \sum_{j=t_{c-2}}^{t_{c+2}} N_{ij} / (t_{c-2} - t_{c+2} + 1)}$$

is the average relative size (measured in number of pups) of source i for the period $[t_{c-2}, t_{c+2}]$.

We specify a Dirichlet ($\alpha_1, \alpha_2, \dots, \alpha_s$) distribution for $\text{Pr}(\mathbf{x})$ with parameters α_i given by $\alpha_i = \alpha_0 E(x_i)$, where $\alpha_0 = \sum_i \alpha_i$ is a parameter that determines the variance of the x_i values and has to be estimated. The prior for $\text{Pr}(\alpha_0)$ is uniform (non-informative) from s to 100, whereas those for $\text{Pr}(w)$ and $\text{Pr}(S)$ are uniform from 0 to 1. The prior for $\text{Pr}(R)$ is uniform from 0 to 5. Having observed the data y (that is, the genotypes in the new populations and the allele frequency distributions of the potential sources), our knowledge about w and $\mathbf{x}$ is given by the posterior distribution

$$\text{Pr}(w, S, R, \alpha_0, \mathbf{x} | y) \propto \text{Pr}(w) \text{Pr}(\alpha_0) \text{Pr}(R) \text{Pr}(\mathbf{x} | S, R, \alpha_0) \text{Pr}(y | w, \mathbf{x})$$

We used the maximum-likelihood estimates of the allele frequencies rather than including them as nuisance parameters, after verifying with synthetic data with the same F_{ST} and sample size that this had negligible effect on the estimation. The posterior distribution was estimated by using the Metropolis–Hastings algorithm^{14,15}.

The DIC scores were produced from MCMC chains of length 520,000 for four models: with R and S unconstrained, with R set to 0, with S set to 0, and with both R and S set to 0. The measure of model complexity was estimated as the difference between the sample mean of the simulated values of the deviance minus an estimate of the deviance by using the simulated values of the parameters θ

$$p_D = \overline{D(\theta)} - (\bar{\theta}).$$

The function $D(\theta)$ is the bayesian deviance given by

$$D(\theta) = -2 \log p(y | \theta) + 2 \log f(y)$$

where $f(y)$ is a standardizing function of the data, y , alone. In our case we used the null standardization obtained by assuming that $f(y)$ is the perfect predictor that gave probability 1 to each observation. DIC was calculated as $p_D + \bar{D}$. The standard error on the DIC scores, estimated from the MCMC chains, was less than 0.040 for each of the estimates.

Received 15 June 2001; accepted 11 January 2002.

- Hanski, I. Metapopulation dynamics. *Nature* **396**, 41–49 (1998).
- Johnson, C. *Insect Migration and Dispersal by Flight* (Methuen, London, 1969).
- Baker, R. *The Evolutionary Ecology of Animal Migration* (Hodder & Stoughton, London, 1978).
- Hanski, I. *Metapopulation Ecology* (Oxford Univ. Press, Oxford, 1999).
- Hiby, A. R., Duck, C. D., Thompson, D., Hall, A. J. & Harwood, J. Seal stocks in Great Britain. *NERC News*, January (1996).
- Pomeroy, P., Twiss, S. & Redman, P. Philopatry, site fidelity and local kin associations within grey seal breeding colonies. *Ethology* **106**, 899–919 (2000).
- Whitlock, M. C. Nonequilibrium population structure in forked fungus beetles: extinction, colonization, and the genetic variance among populations. *Am. Nat.* **139**, 952–970 (1992).
- Ingvarsson, P. K. Kin-structured colonization in *Phalacrocorax substriatus*. *Heredity* **80**, 456–463 (1998).
- Giles, B. E. & Goudet, J. Genetic differentiation in *Silene dioica* metapopulations: Estimation of spatiotemporal effects in a successional plant species. *Am. Nat.* **149**, 507–526 (1997).
- Milner, G., Teel, D., Utter, F. & Winans, G. A genetic method of stock identification in mixed populations of pacific salmon, *Oncorhynchus* spp. *Mar. Fish. Rev.* **47**, 1–8 (1985).
- Rannala, B. & Mountain, J. L. Detecting immigration by using multilocus genotypes. *Proc. Natl Acad. Sci. USA* **94**, 9197–9201 (1997).
- Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B* **64** (in the press).
- Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

The manuscript was read critically by I. Hanksi, R. Lande and P.E. Smouse.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.N. (e-mail: ra.nichols@qmul.ac.uk).

Spatial scale dictates the productivity–biodiversity relationship

Jonathan M. Chase* & Mathew A. Leibold†

* Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

† Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637, USA

The diversity of life is heterogeneously distributed across the Earth. A primary cause for this pattern is the heterogeneity in the amount of energy, or primary productivity (the rate of carbon fixed through photosynthesis), available to the biota in a given location^{1–12}. But the shape of the relationship between productivity and species diversity is highly variable^{10–14}. In many cases, the relationship is 'hump-shaped', where diversity peaks at intermediate productivity^{7,9,10,12,10,15–18}. In other cases, diversity increases linearly with productivity^{4–6,10–12}. A possible reason for this discrepancy is that data are often collected at different spatial scales^{10,12,14}. If the mechanisms that determine species diversity vary with spatial scale, then so would the shape of the productivity–diversity relationship. Here, we present evidence for scale-dependent productivity–diversity patterns in ponds. When the data were viewed at a local scale (among ponds), the relationship was hump-shaped, whereas when the same data were viewed at a regional scale (among watersheds), the relationship was positively linear. This dependence on scale results because dissimilarity in local species composition within regions increased with productivity.

A wide variety of ecological models predict that local-scale species diversity should first increase with slight increases in productivity, but then decline to low levels when productivity is high^{7,9,15,16}. That is, they predict that the productivity–diversity relationship should be 'hump-shaped', and this pattern is often seen in empirical studies at relatively small spatial scales^{7,9,10,12,10,15–18}. At regional spatial scales, considerably less theory is available⁸. However, empirical studies performed at this larger scale often show a very different pattern from those performed at local scales. Instead of being hump-shaped, species diversity often monotonically increases with increasing productivity^{4–6,8–12}. Because studies performed at different spatial scales often consider disparate systems and employ different methodology, it remains unclear as to whether relationships are scale-dependent or whether a single relationship holds across scales.

In this study, we explicitly compared the relationship between primary productivity and species diversity within ponds at two different spatial scales: (1) at the local scale of an individual pond, where local interspecific interactions are likely to have a prominent effect on patterns of species diversity; and (2) at the regional scale of a watershed. We standardized each watershed to include three ponds that were similar in productivity and total area (see Methods). Here, regional processes, such as dispersal among ponds, can interact with local processes to influence patterns of species diversity.

At the local scale (among ponds), both producer and animal species diversity had a statistically significant hump-shaped relationship with primary productivity (Fig. 1a). This result is consistent with other local-scale studies in ponds and lakes^{15,18}. In contrast, at the regional scale (among watersheds), species diversity linearly increased with productivity (Fig. 1b). Thus, we show here that the shape of the productivity–diversity relationship depends on spatial scale.

If the productivity–diversity relationship is scale-dependent, as we observe and as suggested elsewhere^{8–12,14}, then the difference in species composition among localities within regions must increase with productivity. Formally, this can be depicted using the basic diversity equation where regional diversity (γ) is a function of local diversity (α) and species compositional differences among localities (β -diversity) ($\gamma = \alpha\beta$ or $\gamma = \alpha + \beta$)^{19–21}. Given this simple relationship, if the pattern between primary productivity and local (α) diversity is hump-shaped while the relationship between primary productivity and regional (γ) diversity monotonically increases, then species compositional differences among localities (β -diversity) must increase with productivity.

To test this, we calculated species dissimilarity of each watershed by quantifying the species compositional differences among the three ponds that comprise each watershed (see Methods). Although species dissimilarity is not directly analogous to β -diversity, it is conceptually very similar, and the two measures are highly correlated²⁰. It also provides a way of evaluating the difference in species composition among localities without statistically confounding it with estimates of local (α) and regional (γ) diversity^{20,21}.

We found that species dissimilarity increased with productivity (Fig. 2). Ponds within watersheds of low productivity shared the majority of their species, whereas ponds within watersheds of high productivity shared few species. Any individual pond within a high-productivity watershed had relatively few species, but the species composition between ponds was quite different, so that the entire watershed had many species. Thus, the positive correlation between productivity and species dissimilarity (Fig. 2) explains the scale-dependent productivity–diversity relationship (Fig. 1). Further, we suggest that whenever a scale-dependent productivity–diversity relationship is observed in natural ecosystems, the explanation for this pattern will include covariation between productivity and species dissimilarity.

There are three primary mechanisms that would cause species dissimilarity to increase with productivity. First, regions with higher average productivity may also have a larger degree of heterogeneity in environmental factors (including productivity itself), which could increase species dissimilarity. In this study, we did not find a relationship between the variance in productivity among ponds within a watershed and the average productivity of watersheds (linear regression, $P < 0.3$). We also measured a variety of standard physical and chemical variables in these ponds (see Methods). For every variable, we found no relationship between its variance within a watershed and the average productivity of watersheds (for all linear regressions, $P > 0.2$). We conclude that this mechanism is unlikely for this system.

Second, regions with higher average productivity may also have a higher propensity for temporal variation in local species composition, which could increase species dissimilarity. Because of the limited temporal scale of our study (two years), we are unable to evaluate whether this mechanism may have influenced our results.

Third, in some cases, community composition can strongly depend on the order in which species initially entered a community; a phenomenon known as multiple stable states¹⁷. If communities with higher average productivity are more likely to obtain multiple stable states than regions with low average productivity, then species compositional dissimilarity would increase with productivity.

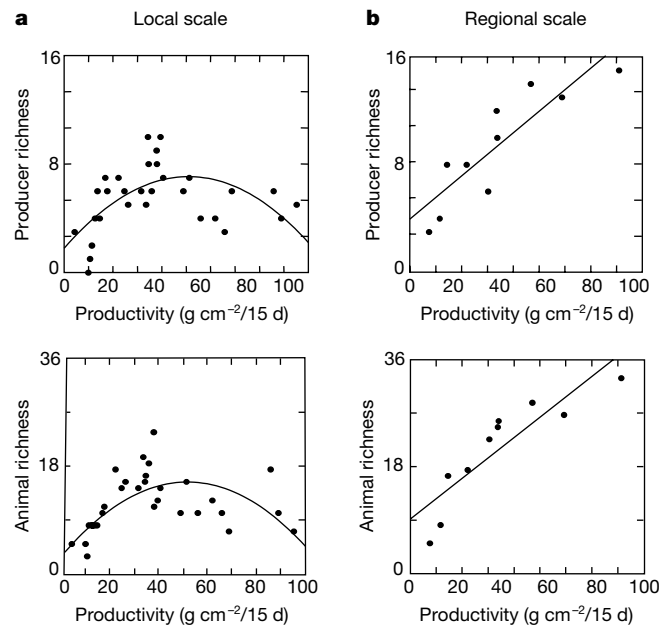


Figure 1 Results from the survey of pond species diversity relative to *in situ* primary productivity at local and regional scales. Top panels, producers (vascular plants and macroalgae); bottom panels, benthic animals (insects, crustaceans, amphibians, and so on). **a**, Local, within-pond, species diversity ($N = 30$). Both relationships are significantly unimodal ($P < 0.05$) (see text for explanation of statistical methodology). The line

represents the estimated quadratic function. **b**, Regional, within-watershed, species diversity. For both producers (regression: $N = 10$, $R^2 = 0.74$, $P < 0.001$) and benthic animals (regression: $N = 10$, $R^2 = 0.75$, $P < 0.001$) regional species diversity was linearly related to primary productivity.

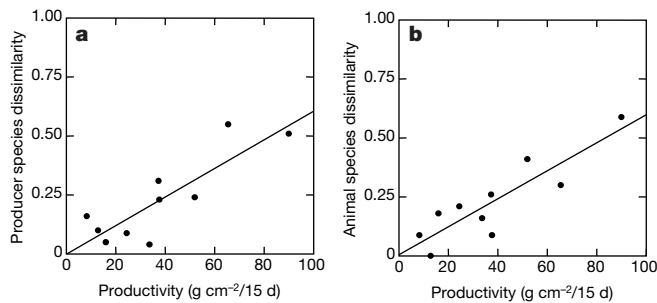


Figure 2 The relative dissimilarity in species composition among local ponds within the watersheds (calculated as one minus Jaccard's index of similarity). Each point represents the average of the dissimilarity estimates from each two-way comparison of species composition among the three ponds within a watershed (see Methods). Regional dissimilarity of both producers **(a)** (regression: $N = 10$, $R^2 = 0.69$, $P < 0.002$), and animals **(b)** (regression: $N = 10$, $R^2 = 0.74$, $P < 0.001$) was linearly positively related to regional productivity.

Again, however, this mechanism cannot be assessed in this study, because the history of these communities is unknown.

We have thus shown how species diversity, when viewed at different spatial scales, can respond in fundamentally different ways to the same environmental factor (productivity in this case). This supports the notion that considerable insight can be gained by increasing the scale, both spatially and temporally, in which species diversity is viewed.

Our results also have implications for understanding how patterns of species diversity might respond to anthropogenic influences. The input levels of nitrogen and phosphorus, important limiting nutrients for productivity, have increased greatly in many ecosystems^{23,24}. This cultural eutrophication often decreases local species diversity^{23,24}. In order to fully understand and manage damaged ecosystems, we must consider the effects of cultural eutrophication on both local and regional species diversity. If, as we found in our study, higher productivity regions contain fewer species locally, but more species regionally, then cultural eutrophication may decrease local species diversity but increase regional species diversity. Alternatively, if cultural eutrophication homogenizes all of the habitats in an area and reduces species dissimilarity, then it can decrease both local and regional species diversity. Finally, species adapted to live in habitats with low productivity could be permanently lost from the ecosystem. Understanding the scale-dependence of these issues will be essential in order to predict and ameliorate the effects of humans on the earth's biota. □

Methods

Estimating primary productivity and other variables

Measuring total nutrient (nitrogen and phosphorus) availability in these ponds does not provide a realistic estimate of primary productivity because ponds also vary considerably in other limiting factors, such as temperature and light availability. Thus, we indirectly estimated primary productivity within each pond by measuring the rate of algal biomass accrual on artificial substrates in the absence of herbivory²⁵. We also estimated several physical and chemical variables of each pond using standard methods²⁵. These variables included pond area, average depth, tree canopy coverage, total nitrogen, total phosphorus, pH, average temperature, and average dissolved oxygen.

Choice of ponds within watersheds

We chose 30 ponds (local) nested within 10 watersheds (regional). To standardize the regional scale, we chose three ponds within each watershed, that occurred within 0.5 km of each other, and that owing to similar environmental conditions and chemical properties of the water and soil, had primary productivity values $\pm 1 \text{ g cm}^{-2}/15 \text{ d}$ of each other (see below). The three chosen ponds within each watershed had a similar total surface area ($\pm 500 \text{ m}^2$), so as not to confound the effect of area on species diversity.

Estimating species diversity

Each pond was sampled twice per year (April–May and July–August) for two years (1997 and 1998), to incorporate both seasonal and year-to-year variation in species composition, for a total of four censuses per pond. Animals (excluding zooplankton) were sampled

using D-net sweeps (0.1-m width, 0.33-mm mesh). Individuals were identified in the field, and sorted to the lowest possible taxonomic category (usually species or in cases where species could not be reliably identified, we categorized them into operational taxonomic units). In each pond, D-net sweeps were made until a full series of ten sweeps revealed no new species. This methodology was used because we could not reliably standardize sampling across ponds that varied in area, and because ponds were highly variable in their bottom substrate and debris, making D-net sweeps difficult to compare directly among ponds.

Producers, including submerged and emergent macrophytes, filamentous macroalgae, and floating duckweeds (but not microscopic periphytic and planktonic algae), were sampled by identifying species along transects through each pond. In each pond, transects were censused until no new species were found per five transects.

For both producers and animals, local species diversity was calculated as the total number of species observed in each pond over all census periods. Regional species diversity was calculated as the total number of species observed in the three ponds within each watershed. We calculated the dissimilarity in species composition among ponds within each watershed (region) as one minus Jaccard's similarity index (if dissimilarity is zero, then all species are shared among local communities; if dissimilarity is one then no species are shared). Because this index is for pairwise comparisons of communities, and our 'regions' consisted of three ponds, we calculated each combination of pairwise dissimilarity values, and averaged them to attain a regional dissimilarity value. Our dissimilarity metric is similar conceptually to β -diversity, but retains slightly different statistical properties²⁰. Nevertheless, dissimilarity is more appropriate for statistical comparisons among regions because β -diversity is confounded owing to its lack of independence from regional species diversity^{20,21}.

Statistical analyses

For each data set, we first performed linear regression to determine whether there was a significant relationship between primary productivity and species diversity at the local or regional scales. We then performed quadratic regressions to see if there were significant deviations from linear relations. When we found significant quadratic effects, we then determined whether it was significantly unimodal (hump-shaped) using a statistical test developed by Mitchell-Olds and Shaw²⁶ (see refs 12, 15, 18 for its usage in the context of productivity–diversity relationships). This method uses quadratic regression to estimate the productivity associated with the peak of the relationship between productivity and diversity, as well as the 95% confidence interval around that peak. Next, the method uses t -tests to determine whether the estimated peak of a data set is significantly greater than species diversity at the minimum and less than species richness at the maximum productivity values.

Received 25 September 2001; accepted 3 January 2002.

1. Lotka, A. J. *Elements of Mathematical Biology* (Dover, New York, 1956).
2. Connell, J. H. & Orias, E. The ecological regulation of species diversity. *Am. Nat.* **98**, 399–414 (1964).
3. MacArthur, R. H. Patterns of species diversity. *Biol. Rev.* **40**, 510–533 (1965).
4. Brown, J. H. Two decades of homage to Santa Rosalia: Towards a general theory of biodiversity. *Am. Zool.* **21**, 877–888 (1981).
5. Currie, D. J. & Paquin, V. Large-scale biogeographical patterns of species richness of trees. *Nature* **329**, 326–327 (1987).
6. Currie, D. J. Energy and large-scale patterns of animal and plant species richness. *Am. Nat.* **137**, 27–49 (1991).
7. Tilman, D. & Pacala, S. in *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds Ricklefs, R. E. & Schluter, D.) 12–25 (Univ. Chicago Press, Chicago, 1993).
8. Wright, D. H., Currie, D. J. & Maurer, B. A. in *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (eds Ricklefs, R. E. & Schluter, D.) 66–76 (Univ. Chicago Press, Chicago, 1993).
9. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
10. Waide, R. B. et al. The relationship between primary productivity and species richness. *Annu. Rev. Ecol. Syst.* **30**, 257–300 (1999).
11. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
12. Mittelbach, G. G. et al. What is the observed relationship between productivity and diversity? *Ecology* **82**, 2381–2396 (2001).
13. Abrams, P. A. Monotonic or unimodal diversity–productivity gradients: What does competition theory predict? *Ecology* **76**, 2019–2027 (1995).
14. Gross, K. L. et al. Patterns of species density and productivity at different spatial scales in herbaceous plant communities. *Oikos* **89**, 417–427 (2000).
15. Leibold, M. A. Biodiversity and nutrient enrichment in pond plankton communities. *Evol. Ecol. Res.* **1**, 73–95 (1999).
16. Ritchie, M. E. & Olff, H. Spatial scaling laws yield a synthetic theory of biodiversity. *Nature* **400**, 557–560 (1999).
17. Abramsky, Z. & Rosenzweig, M. L. Tilman's predicted productivity–diversity relationship shown by desert rodents. *Nature* **309**, 150–151 (1984).
18. Dodson, S. I., Arnott, S. E. & Cottingham, K. L. The relationship in lake communities between primary productivity and species richness. *Ecology* **81**, 2662–2679 (2000).
19. Whittaker, R. H. Evolution and the measurement of species diversity. *Taxon* **21**, 213–251 (1972).
20. Lande, R. Statistics and partitioning of species diversity and similarity among multiple communities. *Oikos* **76**, 393–401 (1996).
21. Loreau, M. Are communities saturated? On the relationship between α , β , and γ diversity. *Ecol. Lett.* **3**, 73–76 (2000).
22. Law, R. in *Advances in Ecological Theory: Principles and Applications* (ed. McGlade, J.) 141–171 (Blackwell, Oxford, 1999).
23. Vitousek, P. M. et al. *Human Alteration of the Global Nitrogen Cycle: Causes and Consequences. Issues in Ecology Vol. 1* (Ecological Society of America, Washington DC, 1997).
24. Carpenter, S. et al. *Nonpoint Pollution of Surface Waters with Phosphorus and Nitrogen. Issues in Ecology Vol. 3* (Ecological Society of America, Washington DC, 1998).

25. Clesceri, L. S., Greeberg, A. E. & Eaton, A. D. (eds) *Standard Methods for the Examination of Water and Wastewater* 20th edn (American Public Health Association, Washington DC, 1998).
26. Mitchell-Olds, T. & Shaw, R. E. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* **41**, 1149–1161 (1987).

Acknowledgements

We thank J. Shurin, M. Willig, M. Vandermuelen, P. Lorch and especially T. Knight for discussions and comments. A. Downing, J. Shurin and T. Leibold helped with the pond survey. This research was supported by the Kellogg Biological Station (Michigan State University), the University of Chicago, the University of California—Davis, the University of Pittsburgh and the NSF.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.C. (e-mail: jchase@pitt.edu).

Correlated binocular activity guides recovery from monocular deprivation

Peter C. Kind^{*,†}, Donald E. Mitchell[‡], Bashir Ahmed^{*}, Colin Blakemore^{*}, Tobias Bonhoeffer[§] & Frank Sengpiel^{||}

^{*} University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK

[†] Department of Biomedical Sciences, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh EH8 9XD, UK

[‡] Department of Psychology, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

[§] Max-Planck-Institut für Neurobiologie, Am Klopferspitz 18a, 82152 München-Martinsried, Germany

^{||} Cardiff School of Biosciences, Museum Avenue, PO Box 911, Cardiff CF10 3US, UK

Monocular deprivation (MD) has much more rapid and severe effects on the ocular dominance of neurons in the primary visual cortex (V1) than does binocular deprivation¹. This finding underlies the widely held hypothesis that the developmental plasticity of ocular dominance reflects competitive interactions for synaptic space between inputs from the two eyes². According to this view, the relative levels of evoked activity in afferents representing the two eyes determine functional changes in response to altered visual experience. However, if the deprived eye of a monocularly deprived kitten is simply reopened, there is substantial physiological and behavioural recovery, leading to the suggestion that absolute activity levels, or some other non-competitive mechanisms, determine the degree of recovery from MD^{3–7}. Here we provide evidence that correlated binocular input is essential for such recovery. Recovery is far less complete if the two eyes are misaligned after a period of MD. This is a powerful demonstration of the importance of cooperative, associative mechanisms in the developing visual cortex.

In normal kittens, artificially induced strabismus, which misaligns the receptive fields of binocular neurons and hence effectively decorrelates the inputs from the two eyes, causes a substantial breakdown of cortical binocularity^{3,8}. We have tested whether strabismus induced immediately after MD prevents physiological and behavioural recovery of the reopened deprived eye. The results demonstrate the role of correlated activity in the recovery process.

Ten kittens were monocularly deprived for 10 days, starting at 5 weeks of age. In five animals, MD was followed by normal binocular experience (MDB). In the other five cats, a convergent strabismus (esotropia) was induced in the nondeprived eye when

the deprived eye was reopened (MDS). In four of these cats, an esotropia of between 10° and 20° persisted throughout the recovery period, as judged from the positions of the corneal reflexes, whereas in one animal (MDS4), the visual axes seemed to become realigned within a few days of the strabismus surgery.

Intrinsic-signal optical imaging of V1 of both hemispheres was performed at least 14 days after the deprived eye had been reopened. In two additional control kittens we performed optical imaging, immediately after a 10-day period of MD, to obtain maps of activation through each eye. In these animals, in agreement with previous results⁹, deprived-eye responses were weak and restricted to small 'islands' that occupied, on average, only 14.3% of cortical territory in V1 ipsilateral to the deprived eye and 18.2% of V1 contralateral to the deprived eye (see Fig. 3a).

All five MDB cats exhibited remarkable recovery of responses through the previously deprived eye; the maps were generally indistinguishable from monocular responses in normal cats of similar age. In the ipsilateral hemisphere (Fig. 1a), on average the previously deprived eye dominated 48.2 ± 3.5% (mean ± s.d.) of the cortical surface, compared with 48.8 ± 3.8% in four normal kittens. In the contralateral hemisphere, the deprived eye's domains occupied 52.6% (±4.2%), compared with 51.2% (±3.8%) in normal kittens.

In contrast, representation of the deprived eye was still much reduced in all four kittens in which inputs from the two eyes were decorrelated because of a persistent strabismus (MDS1–3 and MDS5). In the hemisphere ipsilateral to the deprived eye, responses to stimulation of that eye were often restricted to 'islands' (Fig. 1b; for example, animals MDS2 and MDS3). The deprived eye's territory covered just 36.7% (±3.6%) of the imaged region, significantly less than in the kittens with normal binocular recovery ($P < 0.002$, one-tailed t -test; see Fig. 2). Even if the data were included from the additional animal (MDS4) in which the visual axes became realigned, the mean cortical territory dominated by the deprived eye was 39.4% (±6.8%), significantly smaller than in the five MDB kittens ($P < 0.02$). The results also differed significantly from data for the ipsilateral eye in four kittens of similar age

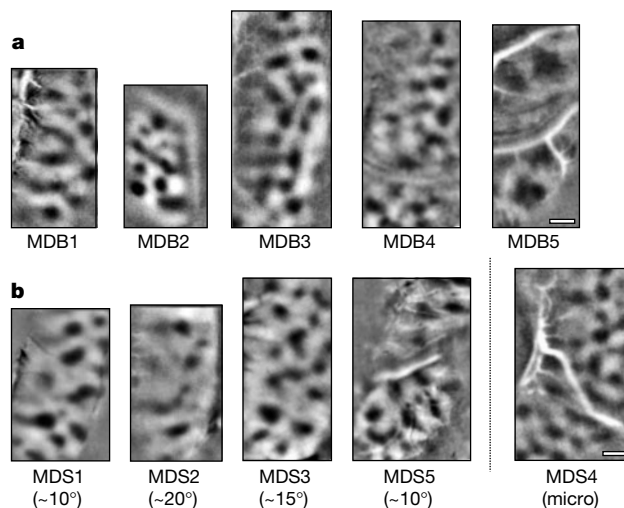


Figure 1 Ocular-dominance maps from left-hemisphere V1 of cats with 10-day MD of the left eye. **a**, Five cats in which MD was followed by concordant binocular vision (MDB); **b**, five cats which were strabismic during the binocular recovery period (MDS). Dark islands represent regions dominated by the left (originally deprived) eye. In all the MDB animals, dark and light areas occupy roughly equal areas (**a**). In the MDS cats (except the microstrabismic cat MDS4 (micro)), dark (left-eye) regions often form relatively isolated patches in a lighter 'sea' of cortex dominated by the right (nondeprived) eye (**b**). Blood vessel artefacts appear as light-grey linear and branching patterns (for example, MDS4 and MDB5). Scale bar, 1 mm. In all images the posterior pole is at the bottom.

made strabismic shortly after eye-opening without prior MD ($48.0 \pm 5.8\%$, $P < 0.01$). For the hemisphere contralateral to the deprived eye, the difference between the strabismic group (MDS1–5, $46.0 \pm 3.1\%$) and the binocular recovery group (MDB1–5) was smaller but also significant ($P < 0.02$, one-tailed t -test; see Fig. 2).

The one kitten in which an obvious strabismus persisted for only a few days after surgical induction (MDS4) served as an interesting control: it had undergone exactly the same procedures as the other strabismic animals but had quickly gained roughly aligned binocular input. The amount of cortical territory dominated by the deprived eye in MDS4 was indistinguishable from that in the MDB kittens (50.2% in the ipsilateral hemisphere and 47.0% in the contralateral hemisphere).

No significant correlation was found between the territory dominated by the deprived eye and the duration of the recovery period in either the MDB group ($r = 0.87$, $P > 0.05$) or the MDS group ($r = 0.05$, $P > 0.9$). This indicates that whatever physiological recovery occurred was complete within 2 weeks. But there was a highly significant negative correlation between the angle of squint and the territory dominated by the deprived eye (averaged across both hemispheres) in the MDS animals ($r = -0.98$, $P < 0.005$).

In addition to reducing the amount of territory recaptured by the previously deprived eye, strabismus also affected the recovery of orientation selectivity. Immediately after MD, activity resulting from deprived-eye stimulation was very weak and restricted to small islands that responded to all stimulus orientations⁹ (see Fig. 3a). However, in the five kittens with normal binocular recovery (MDB), deprived-eye responses to different orientations became spatially well segregated, and normal orientation selectivity was regained (Fig. 3b).

Recovery of orientation selectivity through the previously deprived eye was more variable and weaker overall in the strabismic (MDS) kittens. In some cases, deprived-eye domains responded almost equally well to all orientations (Fig. 3c), as in monocularly deprived control animals. Single-cell recordings from one animal (MDS3) established that this finding was not due to the very close proximity of selective neurons with widely differing orientation preferences. Rather, most neurons driven by the deprived eye exhibited very poor orientation selectivity. Their half-width of

tuning at half-height was $47.7 \pm 5.6^\circ$ (mean $\pm$ s.e.m.; $n = 11$), about twice the value for cells in area 17 of normal cats or cats made strabismic without prior MD¹⁰.

We quantified the orientation selectivity of optical responses through each eye by calculating, for each pixel, an orientation selectivity index, OSI (see Methods), and averaging it across all pixels. The mean OSI through the deprived eye, as a percentage of that through the nondeprived eye, indicates the degree of recovery. Again, the difference between the MDB and MDS groups was greater for the hemisphere ipsilateral to the deprived eye. In the binocular recovery group, the orientation selectivity of responses through the deprived eye was $83.0 \pm 11.3\%$ (mean $\pm$ s.d.) of that through the nondeprived eye. In contrast, in the strabismic group it

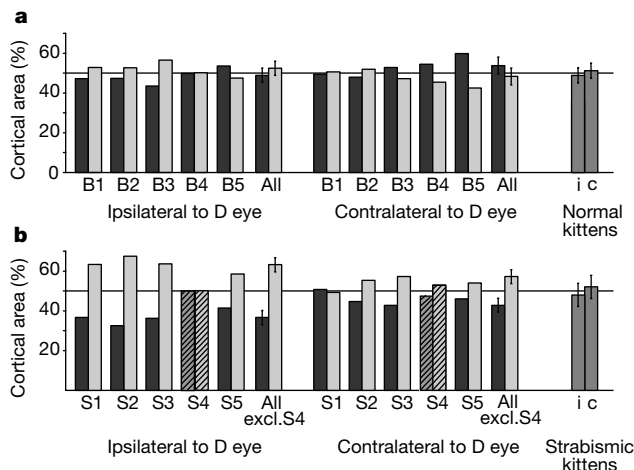


Figure 2 Histograms of cortical surface area dominated by each of the two eyes in MDB cats (a) and in MDS cats (b). Dark bars represent the left, previously deprived eye and light bars the right, non-deprived eye. Data for both hemispheres are displayed separately. In (a), average areas for all five animals are also shown (All), as are results for the ipsilateral (i) and contralateral (c) eyes in four normal kittens. In (b), hatched bars indicate results from kitten MDS4, in which surgical strabismus was transient. Average data from the other four kittens (All excl. S4) are shown, as are results for both eyes in four strabismic kittens.

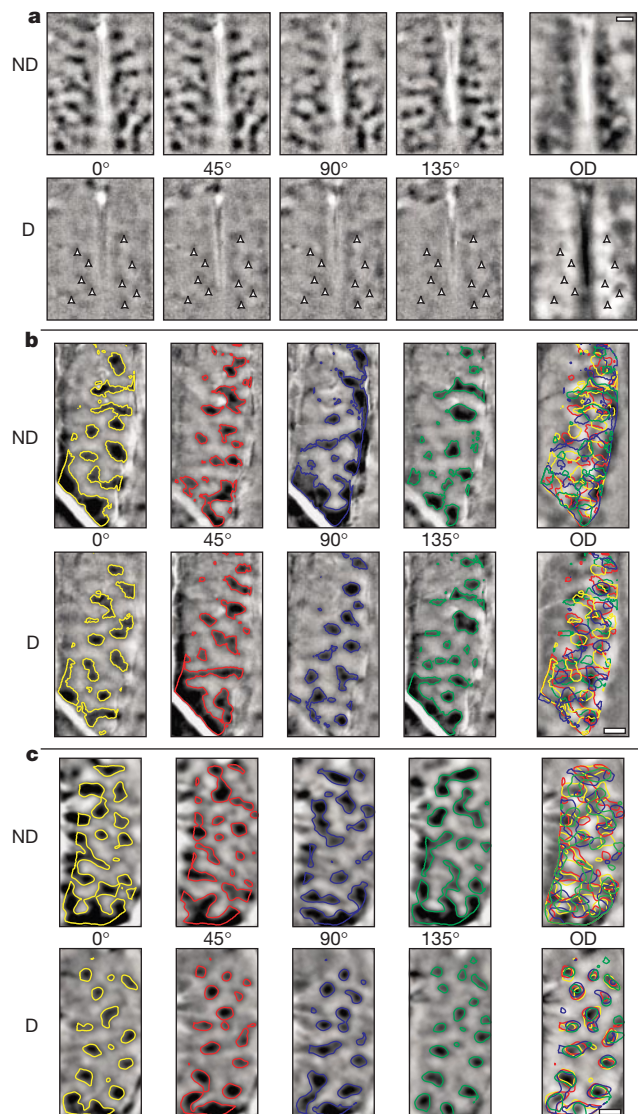


Figure 3 Orientation and ocular-dominance maps for deprived (D) and nondeprived (ND) eyes from a kitten imaged immediately after MD (a), a cat (MDB3) with concordant binocular recovery (b) and a cat (MDS3) with strabismus after MD (c). In a, both hemispheres are shown; in b and c only the left hemisphere is shown. For the control animal (a), arrowheads pointing to small regions still responsive to the deprived eye in the ocular-dominance map have been superimposed on the iso-orientation maps. Iso-amplitude lines in b and c encompass regions of the same minimum signal strength in the four orientation maps. The same lines are then superimposed on the ocular-dominance (OD) maps on the far right. Scale bars, 1 mm; posterior pole is at the bottom.

was only $63.0 \pm 17.4\%$. The difference between the two groups was significant ($P < 0.05$, one-tailed t -test). For the contralateral hemisphere, the relative orientation selectivity through the deprived eye was also greater in the MDB group ($86.6 \pm 6.9\%$) than in the MDS group ($78.1 \pm 18.9\%$) but the difference was not significant. The animal with only transient strabismus (MDS4) showed the greatest recovery of orientation selectivity of all MDS kittens.

Recovery of vision in the deprived eye was tested behaviourally in three littermate pairs of kittens, all subsequently studied by optical imaging. In three MDB kittens, vision through the previously deprived eye recovered faster and attained higher acuity than in three MDS kittens. The MDB kittens began to discriminate gratings of the lowest spatial frequency within 1 or 2 days of the deprived eye being reopened, compared with 3–5 days for MDS animals (Fig. 4). Acuity achieved after 3 weeks was, on average, twice as high in the MDB group (5.44 ± 1.1 cycles deg^{-1} ; mean $\pm$ s.d.) as in the MDS group (2.63 ± 0.13 cycles deg^{-1}). This difference was highly significant ($P < 0.01$, one-tailed t -test). At the same time, acuity in the nondeprived eye did not differ between the two groups (MDB, 7.47 ± 0.37 cycles deg^{-1} ; MDS, 7.03 ± 0.36 cycles deg^{-1} ; $P = 0.11$).

The strabismic animals studied behaviourally included MDS4, whose eyes seemed to become realigned shortly after surgery. Surprisingly, the recovery of acuity in the deprived eye was not superior to that of the other MDS animals, although optical recording revealed substantial re-expansion of the deprived eye's territory. Perhaps realignment of the visual axes was precise enough to provide a correlated input to neurons with relatively large receptive fields, sufficient to account for the optical imaging results, but not to align the smallest receptive fields, which mediate maximal behavioural acuity. The site of the deficit in visual acuity might also be higher in the visual cortical pathway than V1, in agreement with previous reports of reduced visual acuity despite relatively normal neuronal responses in the striate cortex^{11,12}.

Immediately after MD, input from the nondeprived eye dominates the primary visual cortex. However, during a subsequent period of binocular vision there is substantial recovery in the deprived eye. We have now demonstrated that both behavioural and physiological recovery are greatly reduced if the two eyes are misaligned. These results strongly suggest that the correlation of activity between the two eyes promotes recovery through the deprived eye.

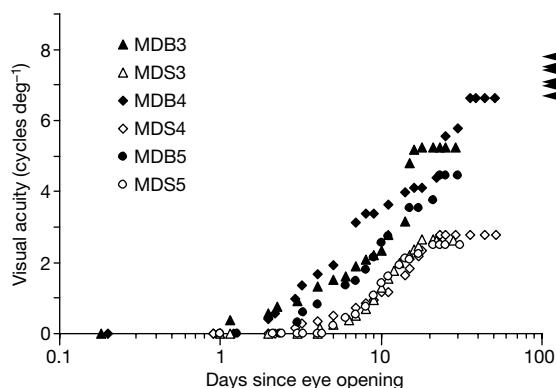


Figure 4 Recovery of visual acuity in the previously deprived eye of three kittens with binocular recovery (MDB, filled symbols) and three kittens with esotropic strabismus (MDS, open symbols) after a period of MD. For littermate pairs of kittens the same type of symbol is used. Grating acuity of the previously deprived eye, obtained with the jumping-stand technique, is plotted against days since the termination of MD. The final acuities of the nondeprived eyes are represented by arrows at the far right. Note that the low acuity through the previously deprived eye cannot reflect a visual deficit associated with squint (strabismic amblyopia), because the esotropia was induced in the nondeprived eye.

If, after MD, the deprived eye is reopened and the other eye is closed (reverse suture), input from the initially deprived eye to the visual cortex recovers completely¹³ (although, interestingly, behavioural recovery is delayed compared with that during simple binocular vision¹⁴). This implies that the absolute level of activity from the initially nondeprived eye alone can drive recovery. The partial recovery in our strabismic animals, compared with that in control kittens examined immediately after MD, might have been due simply to the restoration of activity from the deprived eye. However, some correlated input, sufficient to drive binocular associative mechanisms, might have remained. Surgical strabismus reduces the mobility of the strabismic eye. Hence, during eye movements, the two visual axes would become aligned occasionally, producing correlation of the inputs. Such intermittent correlation would be more frequent for smaller angles of squint. There was indeed a highly significant inverse correlation between the angle of squint and the degree of recovery of input from the deprived eye, as judged by optical imaging, even with angles of deviation above 10° . In fact, intermittent correlation could account for the 20–40% of neurons that maintain binocular input in animals with surgically induced strabismus alone^{8,10}.

At first glance, our findings seem to contradict those of Malach and Van Sluyters¹⁵, who reported that the degree of physiological recovery from MD was similar in kittens with concordant binocular visual input and in those with an induced strabismus in the deprived eye. However, their animals showed only a small strabismic deviation and therefore their greater recovery might, as we suggest for MDS4 in our study, have depended on some degree of correlation of input from the two eyes. Moreover, during the period of MD, Malach and Van Sluyters¹⁵ kept their kittens in complete darkness except for three short (2-h) exposures to light each day. Hence, they were effectively dark-reared for 18 h per day, which might have affected the plasticity of cortical neurons¹⁶.

The need for precise binocular alignment to promote the recovery of input to cortical neurons could account for an otherwise perplexing discrepancy between species. Reopening the deprived eye after MD in monkeys causes little or no physiological recovery unless the nondeprived eye is closed¹⁷—a result that has been taken as evidence for competition. However, the receptive fields in the monkey's foveal representation are so tiny that even a very small misalignment of the visual axes (which frequently occurs after MD^{18–20}) would be likely to compromise the correlation between ocular inputs and hence to prevent recovery.

Our results, together with the effects of reverse suture, therefore argue that both absolute levels of activity⁷ and the degree of correlation of input from the two eyes influence the binocularity of neurons during the postnatal development of the visual cortex. The balance of synaptic input from the two eyes might depend to a large extent on the input from one eye acting as a 'teacher' in an associative learning mechanism for synaptic modification. Many aspects of our findings can be accounted for by the so-called BCM theory of Bienenstock *et al.*²¹, which invokes a changeable 'modification threshold' that depends on integrated synaptic activity. This theory also accounts for the results of many other rearing procedures including MD, reverse suture and dark-rearing^{14,22–25}. □

Methods

Animals

All experimental procedures were performed with the approval of the appropriate British, Canadian and German government authorities. Ten kittens were raised normally until about postnatal day 35 (P35). The left eye was then deprived by lid-suture for 10 d, after which (at P45) it was reopened. For five of the kittens (MDB) the ensuing binocular visual experience was presumably concordant because they had no obvious squint and the projections of their visual axes, determined by direct ophthalmoscopy under paralysis during the optical imaging experiments, were similar to those of normal animals. In the other five animals, esotropia (convergent strabismus) was induced at the time of reopening of the deprived eye by extirpation of the lateral rectus muscle of the other eye (MDS). We chose to make the nondeprived eye strabismic, rather than the deprived eye, so

that any tendency for the induced squint itself to cause a visual impairment (strabismic amblyopia) in the deviating eye would actually reduce the likelihood of the nondeprived eye's continuing to dominate the cortex.

In six kittens (three MDB and three MDS), visual acuity in the previously deprived eye was determined daily by the jumping-stand method^{6,7}. Kittens were trained to discriminate between a vertical and a horizontal grating, the spatial frequency of which was increased in accordance with an ascending method of limits. The nondeprived eye was occluded when the previously deprived eye was tested. The visual acuity of the nondeprived eye was also determined during and after the recovery period, either directly or by measurement of the acuity with both eyes open⁶.

After at least 14 days of recovery, ocular dominance and orientation maps in the primary visual cortex of all ten kittens were obtained by optical imaging of intrinsic signals. For six kittens (three MDB, three MDS), both behavioural tests and optical imaging experiments were performed.

Optical imaging and analysis

Anaesthesia was induced with an intramuscular injection of ketamine (20–40 mg kg⁻¹) and xylazine (2–4 mg kg⁻¹). Animals were intubated and artificially ventilated (50–60% N₂O, 40–50% O₂, 0.9–1.2% halothane). Electrocardiogram, electroencephalogram, end-tidal CO₂ and rectal temperature were monitored continuously. Optical imaging of primary visual cortex was performed as described previously²⁵. Images were captured with either a cooled slow-scan charge-coupled device camera or an enhanced differential imaging system (ORA 2001 or Imager 2001; Optical Imaging Inc.), with the camera focused ~500 μm below the cortical surface. Visual stimuli, produced by a stimulus generator (VSG; Cambridge Research Systems), consisted of high-contrast, sinusoidally modulated gratings (0.2–0.6 cycle deg⁻¹) of four different orientations, drifting at a temporal frequency of 2 Hz, presented to the two eyes separately in randomized sequence, interleaved with trials in which the screen was blank. Single-condition responses (averages of 32–96 trials per eye and orientation) were divided (1) by responses to the blank screen, and (2) by the sum of responses to all four orientations ('cocktail blank')²⁵ to obtain iso-orientation maps. Signal amplitude was displayed on an eight-bit greyscale.

Ocular-dominance maps were obtained by dividing the sum of responses to all four orientations through one eye by the similar sum of responses through the other eye. Resulting maps were high-pass filtered (cutoff 1.3 mm). Within a region of interest that comprised the visually responsive part of the images, excluding blood vessel artefacts, pixels were assigned to the left and right eye, respectively, depending on whether their value was greater or less than 1.

Orientation preference maps were calculated by vectorial addition of four iso-orientation maps, and pseudo-colour coded. Orientation selectivity indices (OSIs) were calculated for responses at each pixel as

$$OSI = \frac{\sqrt{(R_0 - R_{90})^2 + (R_{45} - R_{135})^2}}{R_0 + R_{45} + R_{90} + R_{135}}$$

where R_0 , R_{45} , R_{90} and R_{135} represent the responses in each of the four iso-orientation maps²⁶. OSI represents the magnitude of the orientation preference vector divided by the sum of all responses; it is therefore normalized to values between 0 and 1.

Electrophysiology

In one strabismic animal we determined quantitative orientation–direction tuning curves for 11 single units, recorded with glass-insulated tungsten microelectrodes and discriminated by their spike shapes (Brainware, Oxford, UK). All of these cells were dominated by the previously deprived eye. Responses to drifting gratings (optimum spatial and temporal frequency) of 16 different directions were averaged over five trials of 1.5 s duration. Smooth tuning curves were fitted to the data points on the basis of Fourier analysis²⁷, and preferred orientation and half-width of tuning at half-height were determined from these curves.

Received 9 April 2001; accepted 9 January 2002.

- Wiesel, T. N. & Hubel, D. H. Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.* **28**, 1029–1040 (1965).
- Guillery, R. W. In *The Making of the Nervous System* (eds Parnevelas, J. G., Stern, C. J. & Sterling, R. V.) 356–379 (Oxford University Press, Oxford, 1988).
- Mitchell, D. E. & Timney, B. in *Handbook of Physiology Section I (The Nervous System) Vol. 3 Part 1 (Sensory Processes)* (ed. Darian-Smith, I.) 507–555 (American Physiological Society, Bethesda, 1984).
- Mitchell, D. E., Cynader, M. & Movshon, J. A. Recovery from the effects of monocular deprivation in kittens. *J. Comp. Neurol.* **176**, 53–64 (1977).
- Mitchell, D. E. The extent of visual recovery from early monocular or binocular visual deprivation in kittens. *J. Physiol. (Lond.)* **395**, 639–660 (1988).
- Giffin, F. & Mitchell, D. E. The rate of recovery of vision after early monocular deprivation in kittens. *J. Physiol. (Lond.)* **274**, 511–553 (1978).
- Mitchell, D. E. & Gingras, G. Visual recovery after monocular deprivation is driven by absolute rather than relative visually-evoked activity levels. *Curr. Biol.* **8**, 1179–1182 (1998).
- Hubel, D. H. & Wiesel, T. N. Binocular interaction in striate cortex of kittens reared with artificial squint. *J. Neurophysiol.* **28**, 1041–1059 (1965).
- Crair, M. C., Ruthazer, E. S., Gillespie, D. C. & Stryker, M. P. Relationship between the ocular dominance and orientation maps in visual cortex of monocularly deprived cats. *Neuron* **2**, 307–318 (1997).
- Sengpiel, F., Blakemore, C., Kind, P. C. & Harrad, R. Interocular suppression in the visual cortex of strabismic cats. *J. Neurosci.* **14**, 6855–6871 (1994).
- Kiorpes, L., Kiper, D. C., O'Keefe, L. P., Cavanaugh, J. R. & Movshon, J. A. Neuronal correlates of amblyopia in the visual cortex of macaque monkeys with experimental strabismus and anisometropia. *J. Neurosci.* **18**, 6411–6424 (1998).

- Murphy, K. M. & Mitchell, D. E. Bilateral amblyopia following a short period of reverse occlusion in kittens. *Nature* **323**, 536–538 (1986).
- Blakemore, C. & Van Sluyters, R. C. Reversal of the physiological effects of monocular deprivation in the kitten: further evidence for a sensitive period. *J. Physiol. (Lond.)* **237**, 195–216 (1974).
- Mitchell, D. E., Gingras, G. & Kind, P. C. Initial recovery of vision after early monocular deprivation in kittens is faster when both eyes are open. *Proc. Natl Acad. Sci. USA* **98**, 11662–11667 (2001).
- Malach, R. & Van Sluyters, R. C. Strabismus does not prevent recovery from monocular deprivation: a challenge for simple Hebbian models of synaptic modification. *Vis. Neurosci.* **3**, 267–273 (1989).
- Kirkwood, A., Lee, H.-K. & Bear, M. F. Co-regulation of long-term potentiation and experience-dependent synaptic plasticity in visual cortex by age and experience. *Nature* **375**, 328–331 (1995).
- Blakemore, C., Vital-Durand, F. & Garey, L. J. Recovery from monocular deprivation in the monkey. I. Recovery of physiological effects in the visual cortex. *Proc. R. Soc. Lond. B* **213**, 399–423 (1981).
- Blake, R., Crawford, M. L. & Hirsch, H. V. Consequences of alternating monocular deprivation on eye alignment and convergence in cats. *Invest. Ophthalmol.* **13**, 121–126 (1974).
- Movshon, J. A. Reversal of the behavioural effects of monocular deprivation in the kitten. *J. Physiol. (Lond.)* **261**, 175–187 (1976).
- Quick, M. W., Tigges, M., Gammon, J. A. & Boothe, R. G. Early abnormal visual experience induces strabismus in infant monkeys. *Invest. Ophthalmol. Vis. Sci.* **30**, 1012–1017 (1989).
- Bienenstock, E. L., Cooper, L. N. & Munro, P. W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**, 32–48 (1982).
- Clothetaux, E. E., Bear, M. F. & Cooper, L. N. Synaptic plasticity in visual cortex: comparison of theory with experiment. *J. Neurophysiol.* **66**, 1785–1804 (1991).
- Rittenhouse, C. D., Shouval, H. Z., Paradiso, M. A. & Bear, M. F. Monocular deprivation induces homosynaptic long-term depression in visual cortex. *Nature* **397**, 347–350 (1999).
- Kind, P. C. Cortical plasticity: is it time for a change? *Curr. Biol.* **9**, R640–R643 (1999).
- Bonhoeffer, T. & Grinvald, A. in *Brain Mapping: The Methods* (eds Toga, A. W. & Mazziotta, J. C.) 55–97 (Academic, London, 1996).
- Bonhoeffer, T., Kim, D. S., Maloney, D., Shoham, D. & Grinvald, A. Optical imaging of the layout of functional domains in area 17 and across the area 17/18 border in cat visual cortex. *Eur. J. Neurosci.* **7**, 1973–1988 (1995).
- Wörgötter, F. & Eysel, U. T. Quantitative determination of orientational and directional components in the response of visual cortical cells to moving stimuli. *Biol. Cybern.* **57**, 349–355 (1987).

Acknowledgements

We thank M. Bear and H. Shouval for their helpful discussions, and I. Kehr and M. Jones for technical assistance. This work was supported by the Wellcome Trust (P.C.K.), the Medical Research Council (C.B., F.S., B.A.), Max-Planck-Gesellschaft (T.B., F.S.), the Canadian Institutes of Health Research (D.E.M.) and the Oxford McDonnell Centre for Cognitive Neuroscience (C.B., B.A.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C.K. (e-mail: pkind@ed.ac.uk).

Spike-timing-dependent synaptic modification induced by natural spike trains

Robert C. Froemke & Yang Dan

Division of Neurobiology, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

The strength of the connection between two neurons can be modified by activity, in a way that depends on the timing of neuronal firing on either side of the synapse^{1–10}. This spike-timing-dependent plasticity (STDP) has been studied by systematically varying the intervals between pre- and postsynaptic spikes. Here we studied how STDP operates in the context of more natural spike trains. We found that in visual cortical slices the contribution of each pre-/postsynaptic spike pair to synaptic modification depends not only on the interval between the pair, but also on the timing of preceding spikes. The efficacy of each spike in synaptic modification was suppressed by the preceding spike in the same neuron, occurring within several tens of

milliseconds. The direction and magnitude of synaptic modifications induced by spike patterns recorded *in vivo* in response to natural visual stimuli were well predicted by incorporating the suppressive inter-spike interaction within each neuron. Thus, activity-induced synaptic modification depends not only on the relative spike timing between the neurons, but also on the spiking pattern within each neuron. For natural spike trains, the timing of the first spike in each burst is dominant in synaptic modification.

Whole-cell recordings were made from pyramidal neurons in layer 2/3 (L2/3) of rat visual cortical slices to monitor excitatory postsynaptic potentials (EPSPs) evoked by extracellular stimulation applied in the same layer. To understand how complex spike trains induce synaptic modification, we first characterized the dependence of synaptic modification on the interval between the pre- and postsynaptic spikes, using a standard pairing protocol^{6,8–10}. Each pairing consisted of a single-pulse presynaptic stimulation and a brief postsynaptic depolarizing current injection that induced an action potential. After 60–80 pairings (0.2 Hz) at positive intervals

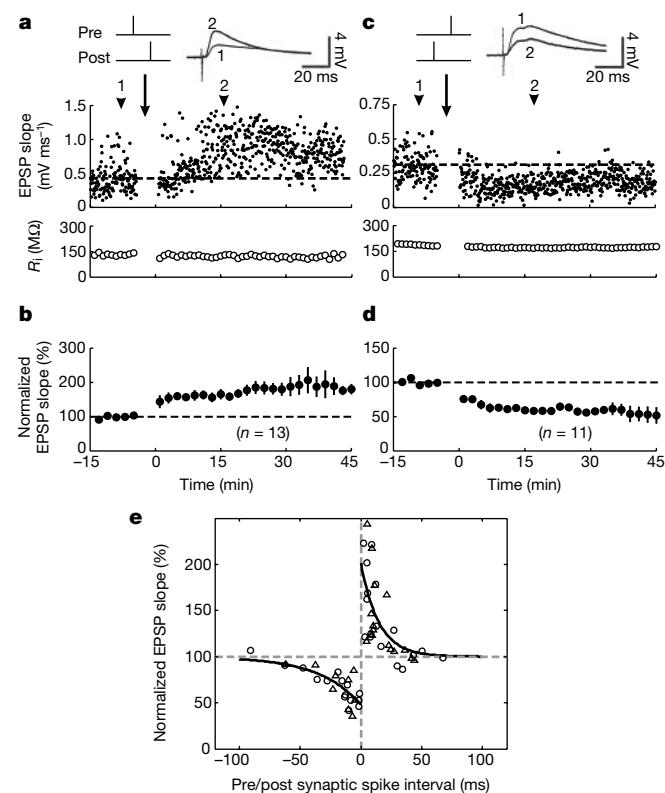


Figure 1 Synaptic modification of L2/3 visual cortical connections induced by pre-/postsynaptic spike pairs. **a**, Example of LTP (117%) induced by repetitive pre- and postsynaptic spiking at a positive interval (9.1 ms). Arrow, induction. Bottom, input resistance (R_i). **b**, Summary of effects of pre → post spiking (increase of $62.5 \pm 12\%$, $n = 13$, $P < 0.001$, t -test; intervals: 2–15 ms). **c**, As in **a**, except LTD (-48%) was induced by a post → pre spike pair (interval, -2 ms). **d**, Summary of effects of post → pre spiking ($-41.5 \pm 4\%$, $n = 11$, $P < 0.0001$, intervals: -2 to -15 ms). **e**, Dependence of synaptic modification on pre/post inter-spike interval. Each point represents one experiment. Circles, normal ACSF ($A_+ = 103 \pm 10\%$, s.d., non-parametric bootstrap, $\tau_+ = 13.3 \pm 1.7$ ms, $n = 17$; $A_- = -51 \pm 1\%$, $\tau_- = 34.5 \pm 1.6$ ms, $n = 15$). Triangles, high divalent ACSF with bicuculline ($A_+ = 102 \pm 11\%$, $\tau_+ = 15.5 \pm 3.2$ ms, $n = 15$; $A_- = -52 \pm 6\%$, $\tau_- = 33.2 \pm 5.3$ ms, $n = 9$). No significant difference between parameters for the two solutions ($P > 0.5$). Curves, single-exponential, least-square fits of the combined data. Mean error of the fit, $18.2 \pm 2.1\%$. Correlation coefficient between the data and the fit, 0.88 . $R^2 = 1 - \sum e_i^2 / \sum y_i^2$ (where e_i is the error, y_i is the measured effect of the i th experiment) = 0.72 .

(pre → post) of 2 to 15 ms, we observed long-term potentiation (LTP) (Fig. 1a, b). The same number of pairings at negative intervals (post → pre, -2 to -15 ms), however, induced long-term depression (LTD) (Fig. 1c, d). Figure 1e summarizes the observed synaptic modification as a function of the pre/post inter-spike interval. Synaptic potentiation was observed at intervals between 0 and 20 ms, whereas depression was observed between 0 and -40 ms, comparable to the temporal window for STDP found at several other glutamatergic excitatory synapses^{5,6,8–9}. Similar LTP and LTD windows were also observed in experiments performed in high divalent external solution containing 4 mM Mg²⁺, 4 mM Ca²⁺ (to reduce polysynaptic transmission), and 3 μ M bicuculline (antagonist of the GABA_A receptor) (Fig. 1e, triangles). Thus the observed STDP is independent of polysynaptic transmission and cortical inhibition. To obtain a quantitative description of the temporal window, we fitted the data on each side with an exponential function $\Delta w = Ae^{-|\Delta t|/\tau}$, where Δw is the percentage change in synaptic strength, A and τ are the scaling factor and time constant of the exponential function, respectively, and Δt is the pre/post inter-spike interval. A and τ were found to be 101% and 14.8 ms respectively for potentiation, and -52% and 33.8 ms for depression.

To predict the effect of a pair of complex spike trains in synaptic modification, a straightforward approach is to combine the contri-

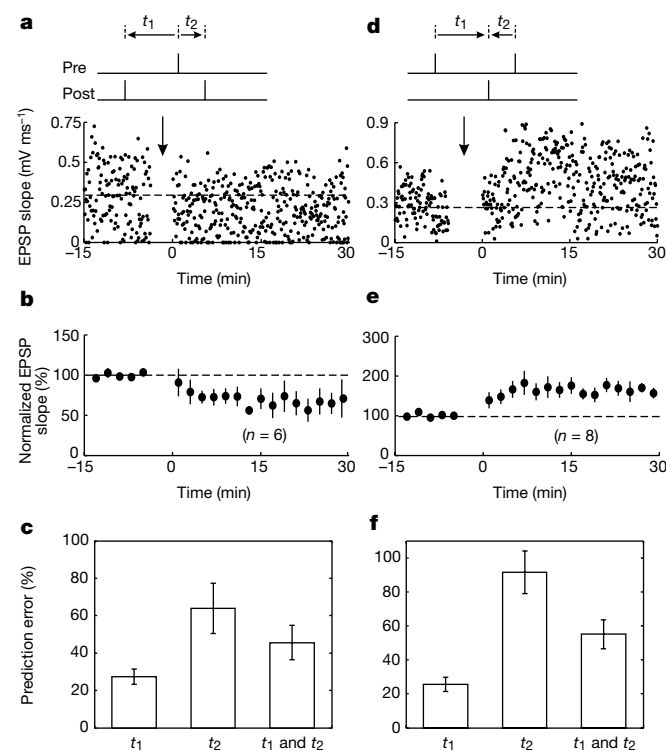


Figure 2 Synaptic modification induced by spike triplets. **a**, Example of LTD induced by a '1/2' triplet. Pre/post spike pair interval defined as $t_2^{\text{post}} - t_1^{\text{pre}}$. Right arrows, positive; left arrows, negative. **b**, Summary of '1/2' experiments satisfying: (1) $t_1 < 0$, (2) $t_2 > 0$, (3) $|t_1 - t_2| \leq 30$ ms, and (4) prediction of the independent model was potentiation or no change. The measured effect was depression ($-33 \pm 11\%$, $n = 6$, $P < 0.05$). **c**, Mean prediction errors based on t_1 , t_2 , and t_1 and t_2 combined (using the independent model) for all the '1/2' triplets such that both $|t_1|$ and $|t_2| < 15$ ms ($n = 11$). Error bars, $\pm$ s.e.m. **d**, As in **a**, except LTP was induced by a '2/1' triplet. **e**, Summary of '2/1' triplet experiments satisfying: (1) $t_1 > 0$, (2) $t_2 < 0$, (3) $|t_1 - t_2| \leq 30$ ms, and (4) prediction of the independent model was depression or no change. The measured effect was potentiation ($65.0 \pm 17\%$, $n = 8$, $P < 0.01$). **f**, Mean prediction errors based on t_1 , t_2 , and t_1 and t_2 combined (using the independent model) for all '2/1' triplets such that both $|t_1|$, $|t_2| < 15$ ms ($n = 14$). Prediction error based on t_1 was smaller than that based on t_2 ($P < 0.001$) and t_1 and t_2 combined ($P < 0.005$).

butions of all pre/post spike pairs in the two spike trains^{11–13}. However, the contribution of each pre/post spike pair may depend not only on its interval, as shown in Fig. 1e, but also on the presence of other spikes in both neurons in an unknown manner. Thus we carried out a series of experiments, using spike patterns of increasing complexity, to examine whether the contribution of each pre/post spike pair depends on the presence of other spikes in the pre- and postsynaptic cells and to develop a quantitative method for estimating the effects of complex spike trains on synaptic strength.

We first added a third spike either pre- or postsynaptically to form a 'triplet', each repeated 60–80 times at 0.2 Hz to induce synaptic modification. A typical experiment using a '1/2' triplet (1 pre and 2 post) is shown in Fig. 2a. According to the temporal window measured with isolated pre/post spike pairs (Fig. 1e), the first post → pre spike pair (spike time, $t_1 = -24$ ms) was expected to reduce the synaptic strength by 26%, and the following pre → post spike pair ($t_2 = 6$ ms) should increase the strength by 67%. Assuming that the two spike pairs contribute independently to synaptic modification and that the contributions are combined multiplicatively (see Methods), this triplet should induce a 24% synaptic potentiation. However, we observed a clear reduction (32%) of synaptic strength. Similar results were obtained in all six experiments using such triplets (Fig. 2b). To characterize the contribution of each spike pair in synaptic modification, we further measured the effects of '1/2' triplets over a range of t_1 and t_2 values ($n = 25$), and compared the results with the predicted synaptic modification based on t_1 , t_2 , or t_1 and t_2 combined (assuming independent contributions of the two spike pairs). We found that when the three spikes were in close proximity (both $|t_1|$ and $|t_2| < 15$ ms), the prediction based on t_1 was significantly better than that based on t_2 (Fig. 2c, $P < 0.01$, Student's paired t -test), indicating that the first spike pair played a dominant role in synaptic modification. Moreover, the prediction based on t_1 was also better than that based on t_1 and t_2 combined (Fig. 2c). This further indicated that the two spike pairs did not contribute independently to synaptic modification, and the contribution of the second pair was strongly suppressed by the presence of the preceding postsynaptic spike. This suppressive effect, however, depended on the proximity of the preceding spike. For $t_1 < -30$ ms, the synaptic modification induced by the triplet could be largely predicted by t_2 (see Fig. 3b).

In a complementary series of experiments, we measured the effects of '2/1' triplets (2 pre, 1 post) to determine whether the contribution of a pre/post spike pair can also be affected by a preceding presynaptic spike. In the example shown in Fig. 2d, the first pre → post spike pair ($t_1 = 6.5$ ms) should induce a 65% potentiation, whereas the following post → pre spike pair ($t_2 = -0.5$ ms) should cause a 52% reduction of synaptic strength. Assuming independent contributions of the two spike pairs (combined multiplicatively), the expected effect of the triplet is a 20% reduction of synaptic strength. The observed effect, however, was a strong synaptic potentiation of 85%. Similar results were obtained in all eight experiments using such triplets (Fig. 2e), again indicating a dominant contribution of the first spike pair. When both t_1 and t_2 were within ± 15 ms, the effect of the triplet was best predicted by t_1 (Fig. 2f). For $t_1 > 30$ ms, however, the synaptic modification induced by the triplet could be largely predicted by t_2 (see Fig. 3b). Thus, a preceding spike in the presynaptic neuron can also suppress the contribution of a pre/post spike pair in synaptic modification.

On the basis of the above observations we proposed a simple model for the inter-spike interactions in synaptic modification. The contribution of each pre/post spike pair depends not only on the interval between the pair (Fig. 1e), but also on the spike 'efficacy', which is suppressed by the preceding spike in each neuron. The spike efficacy is reduced to zero immediately after the preceding spike, and recovers exponentially towards one (Fig. 3a). By fitting

the data of '2/1' triplets ($n = 28$), the time constant of suppression between consecutive presynaptic spikes was found to be 34 ms. Similarly, from the data of '1/2' triplets ($n = 25$), the time constant of suppression in the postsynaptic neuron was found to be 75 ms. To evaluate the validity of the suppression model in accounting for the effects of spike triplets, we compared the predicted and measured effects of both '1/2' and '2/1' triplets within a window of ± 100 ms for t_1 and t_2 . We found that the model correctly predicted the

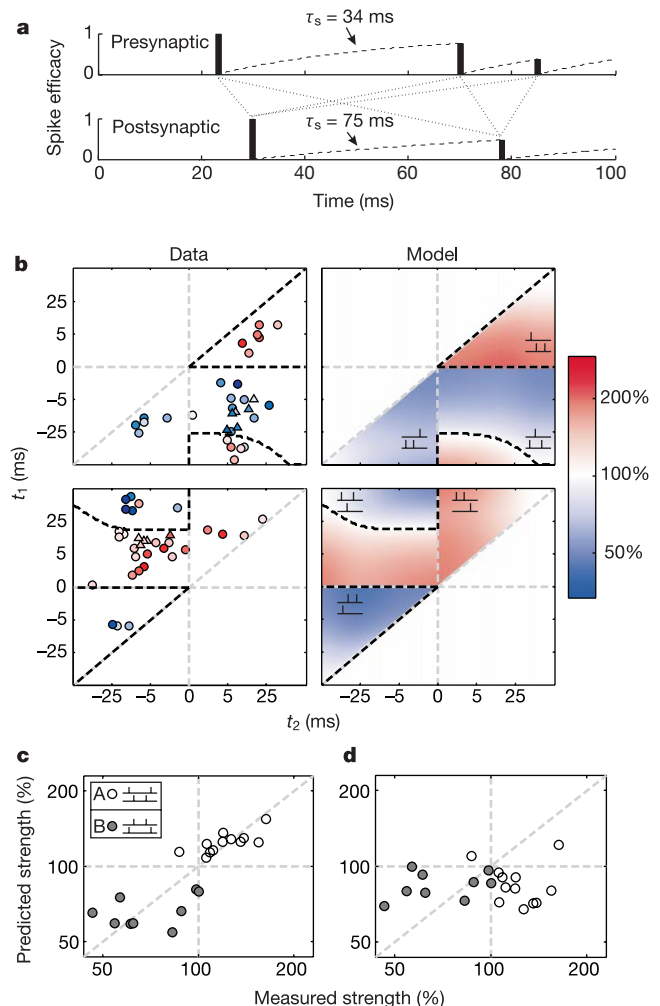


Figure 3 Suppressive interaction between consecutive spikes for triplets and quadruplets. **a**, Model of inter-spike suppression in pre- and postsynaptic cells. Vertical lines indicate spikes, with height indicating efficacy. Dashed curves, efficacy as a function of time. Dotted lines, pre/post spike pairs. **b**, Comparison between measured (Data) and predicted (Model) synaptic modification induced by '1/2' (top) and '2/1' triplets (bottom). Circles, normal ACSF; triangles, high divalent ACSF with bicuculline; each symbol represents one experiment. Dashed lines, borders between predicted potentiation and depression regions. Scale, degree of synaptic modification (red, potentiation; blue, depression). Suppression model: correlation coefficients between predicted and measured effects are 0.89 for '1/2', 0.76 for '2/1' triplets; R^2 is 0.78 for '1/2', 0.59 for '2/1' triplets. Independent model: correlation coefficients are 0.74 for '1/2', 0.55 for '2/1' triplets; R^2 is 0.19 for '1/2', 0.02 for '2/1' triplets. **c**, Predicted versus measured effects of spike quadruplets, based on the suppression model. Open symbols, type A, pre → post → post → pre ($n = 12$, intervals: pre → post, 8.8 ± 2.0 ms, s.d.; post → pre, -9.6 ± 5.2 ms; post → post, 10.6 – 181.8 ms). Filled symbols, type B, post → pre → post → post ($n = 9$, intervals: pre → post, 9.0 ± 6.0 ms; post → pre, -7.9 ± 2.3 ms; pre → pre, 9.6 – 102.5 ms). Mean prediction error, $14.6 \pm 2.6\%$; correlation coefficient, 0.85; R^2 , 0.72. **d**, As in **c**, except predictions are based on the independent model. Mean prediction error, $33.7 \pm 4.5\%$, larger than the suppression model ($P < 0.005$); correlation coefficient, 0.06; R^2 , -0.24 .

direction of synaptic modification in 50/53 triplet experiments (Fig. 3b), suggesting that the suppression model with only two free parameters provides a good description of synaptic modification induced by spike triplets.

In principle, the inter-spike suppression may be mediated by the inhibitory synaptic interactions in the local cortical circuitry¹⁴. However, significant suppression was also observed in both '1/2' and '2/1' triplet experiments in high divalent solution containing bicuculline (Fig. 3b, triangles), indicating that the suppression is mediated by mechanisms in the pre- or postsynaptic neurons rather than by the polysynaptic cortical circuitry. Consistent with previous studies^{15,16}, we noticed significant paired-pulse depression at these intracortical synapses (data not shown), suggesting that the suppression between presynaptic spikes during synaptic modification may result from short-term depression of transmitter release¹⁷ or desensitization/saturation of postsynaptic glutamate receptors¹⁸. The suppression between postsynaptic spikes, on the other hand, may be due to Ca^{2+} -dependent conductances¹⁹ or downstream mechanisms involved in activity-dependent synaptic modification^{20,21}. When more quantitative information concerning these processes becomes available, detailed biophysical models^{22,23} may be constructed to describe the inter-spike interactions observed here. In the following experiments we examined whether the simple suppression model derived from the triplet experiments provides a useful description of the effects of complex spike patterns in synaptic modification.

To extend the complexity of the spike pattern, we next added another spike to form 'quadruplets' (2 pre, 2 post), with the spike sequence being either pre $\rightarrow$ post $\rightarrow$ post $\rightarrow$ pre (type A) or post $\rightarrow$ pre $\rightarrow$ pre $\rightarrow$ post (type B) (Fig. 3c, inset). These patterns con-

tained two pre/post spike pairs with positive intervals and two pairs with negative intervals, and the measured effects of these quadruplets provide a direct test of the inter-spike suppression within each neuron. According to the suppression model, as the first spike of each neuron plays a dominant role in determining the sign and magnitude of synaptic modification, the type-A quadruplets are likely to induce potentiation whereas type-B quadruplets are likely to induce depression. However, without suppression the two types of quadruplets should have similar effects. As shown in Fig. 3c, the measured effects of the two types of quadruplets were clearly different, with type A (open circles) inducing mostly potentiation and type B inducing depression (filled circles). The measured effects agreed well with the suppression model, but were inconsistent with the prediction based on independent contributions of the spike pairs (Fig. 3d).

Finally, we examined synaptic modification in cortical slices induced by natural spike patterns occurring *in vivo* in response to visual stimuli. Spike trains were recorded simultaneously from neighbouring V1 neurons of the anaesthetized cat in response to time-varying natural scenes (Fig. 4a). Segments of spike trains (1 s, containing 6–12 spikes) from two neurons with partially overlapping receptive fields were applied to pre- and postsynaptic neurons in the slice to induce synaptic modification. Figure 4b shows three examples of these spike train segments, each of which induced a significant long-term synaptic modification after 60 repetitions (0.2 Hz). Figure 4c summarizes the results of 22 experiments using different pairs of natural spike train segments. We found that the observed synaptic modification agreed well with that predicted by the suppression model. In contrast, the data fit poorly with the prediction based on independent contributions of all spike

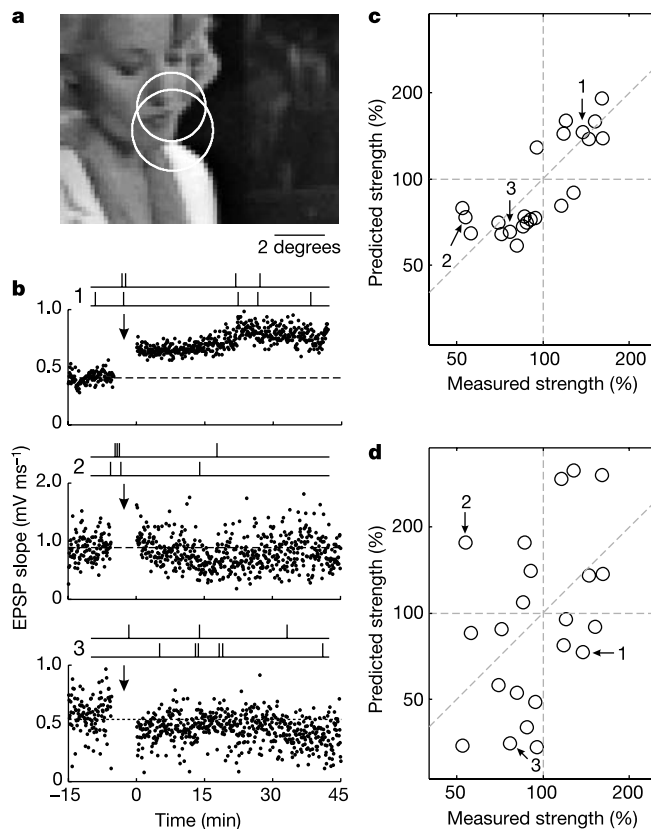


Figure 4 Synaptic modification induced by natural spike-train segments. **a**, Frame (64×48 pixels, 0.17° per pixel) in a movie used to evoke cortical responses in the cat. Circles, receptive fields. **b**, Examples of synaptic modification induced by pairs of natural spike-train segments (1 s). Synaptic modification: top, 37%; middle, -47%; bottom, -23%. **c**, Predicted versus measured effect, based on the suppression model. Arrows,

experiments in **b**. Mean prediction error, $20.7 \pm 2.4\%$ (s.e.m., $n = 22$); correlation coefficient, 0.79; R^2 , 0.53. **d**, As in **c**, but predictions are based on the independent model. Mean prediction error, $54.8 \pm 6.6\%$, larger than the suppression model ($P < 0.0001$); correlation coefficient: 0.42; R^2 , -2.33.

pairs (Fig. 4d). Thus, the dependence on the pre/post inter-spike interval (Fig. 1e) and the suppressive interaction between consecutive spikes within each neuron (Fig. 3) together provide a good description of STDP induced by complex spike patterns encountered *in vivo*.

Spike-timing-dependent modification of excitatory connections between L2/3 neurons is likely to play an important role in plasticity of visual cortical circuitry *in vivo*. Recent studies have shown that timing of visual stimuli on the order of 20 ms can affect the direction and magnitude of plasticity in orientation tuning of adult cortical neurons²⁴, and such functional plasticity may be mediated by STDP of excitatory intracortical connections. A direct implication of the inter-spike suppression shown here is that timing of the first spike in each burst has a crucial role in determining the sign and magnitude of synaptic modification. In many cortical neurons, stimulus onset or saccadic eye movement evokes fast initial transients followed by sustained responses, and our results suggest that timing of the initial transients may be especially important for long-term synaptic modification. Previous studies have shown that presynaptic spike timing in complex spike trains is important in neural signalling owing to short-term synaptic plasticity^{15,16,25}. Studies of STDP have also underscored the critical role of precise spike timing in long-term synaptic modification^{3–10}. Our results show that the spike timing dependence of long-term synaptic modification induced by complex spike trains must include not only the relative timing between pre- and postsynaptic spiking, but also the inter-spike intervals within each neuron. □

Methods

Visual cortical slice preparation

Acute visual cortical slices were prepared from 2–5-week-old Sprague–Dawley rats (no correlation between age and synaptic modification; LTP: $r = 0.09$, $P > 0.4$, $n = 67$; LTD: $r = -0.16$, $P > 0.2$, $n = 62$; analysis of variance, ANOVA). Rats were deeply anaesthetized with halothane, decapitated, and the brain quickly placed into ice-cold dissection buffer containing (in mM): 206 sucrose, 2 KCl, 2 MgSO₄, 1.25 NaH₂PO₄, 1 CaCl₂, 1 MgCl₂, 26 NaHCO₃, and 10 dextrose, bubbled with 95% O₂/5% CO₂ (pH 7.4). Coronal visual cortical slices (300–400 μm thick) were prepared with a vibratome (Pelco), placed in warm dissection buffer (33–35 °C) for <30 min, transferred to a holding chamber containing artificial cerebrospinal fluid (ACSF, in mM: 124 NaCl, 2 KCl, 1.5 MgSO₄, 1.25 NaH₂PO₄, 2.5 CaCl₂, 26 NaHCO₃ and 10 dextrose), and kept at 22–24 °C for >1 h before use. For experiments, slices were transferred to the recording chamber and perfused (4.0–4.5 ml min⁻¹) with oxygenated ACSF at 22–24 °C. Some specified experiments were performed in modified ACSF containing 4 mM CaCl₂, 4 mM MgSO₄, and 3 μM bicuculline methiodide. Several experiments performed at higher temperature (35.8 ± 1.5 °C) confirmed the dependence of synaptic modification on both the pre/post and inter-spike intervals within each neuron (data not shown).

Electrophysiology

Somatic whole-cell recordings were made in current-clamp with an Axopatch 200B amplifier (Axon) using infrared differential interference optics (IR-DIC) video microscopy. L2/3 pyramidal cells were selected on morphology and regular spiking in response to current injection²⁶. Patch pipettes (3–7 MΩ) were filled with intracellular solution (in mM: 120 K-gluconate, 10 HEPES, 0.1 EGTA, 20 KCl, 2 MgCl₂, 10 phosphocreatine, 2 ATP, and 0.25 GTP). The mean resting potential was -71.1 ± 0.8 mV, corrected for the measured liquid junction potential (6.8 mV). The series resistance was 13.3 ± 0.8 MΩ. R_i (153.2 ± 7.8 MΩ) was monitored with hyperpolarizing current pulses (50 pA, 100 ms); cells were excluded if input resistance R_i changed >30% over the entire experiment. Data were filtered at 2 kHz, digitized at 10 kHz, and analysed with Clampfit 8 (Axon). Extracellular stimulation (0.1–1 ms, 5–150 μA) was applied in L2/3 with a small glass bipolar electrode 0.05–1.0 mm from the recording electrode. EPSP size: 3.2 ± 2.8 mV (s.d., $n = 100$). The initial slope of EPSP (first 2 ms) was used to calculate synaptic strength, as this component reflects monosynaptic input^{8,10}. To ensure L2/3 pathway specificity, we placed a second stimulating electrode in L4 in the same column and monitored the evoked EPSPs. During induction, L4 stimulation was temporarily interrupted. In experiments where synaptic modification was induced in L2/3, the L4 EPSPs were unaffected, indicating independence of L2/3 and L4 stimulation. Stable baselines of synaptic strength were established by 6–12 min of stimulation at 0.2 Hz. Synaptic strength after induction was measured 11–20 min after the end of the induction protocol. During induction, postsynaptic spiking was evoked with depolarizing current pulses (1 nA, 2–3 ms). Presynaptic spike timing was defined as the onset of EPSP and postsynaptic spike timing was measured at the peak of the action potential^{5,6,8}. To collect natural spike patterns, natural scene movies were used as visual stimuli, and extracellular recordings were made in cat visual cortex^{24,27}. Spike train segments for slice experiments were sampled from five simultaneously recorded cell pairs.

Predicting the effects of spike train segments

To predict the effects of spike train segments using the suppression model, each pre- and postsynaptic spike was assigned an efficacy, which depends only on the interval from the preceding spike in the same neuron: $\epsilon_i = 1 - e^{-(t_i - t_{i-1})/\tau_s}$, where ϵ_i is the efficacy of the i th spike, t_i and t_{i-1} are the timings of the i th and $(i-1)$ th spike, respectively, and τ_s is the suppression time constant. The contribution of each pre/post spike pair to synaptic modification was estimated as $\Delta w_{ij} = \epsilon_i^{\text{pre}} \epsilon_j^{\text{post}} F(\Delta t_{ij})$, where Δw_{ij} is the synaptic modification due to the i th presynaptic spike and the j th postsynaptic spike, ϵ_i^{pre} and ϵ_j^{post} are the efficacies of the two spikes, respectively, and Δt_{ij} is the interval between the two spikes, $t_j^{\text{post}} - t_i^{\text{pre}}$. The function F represents the temporal window for STDP measured with isolated spike pairs (Fig. 1e), expressed as:

$$F(\Delta t) = \begin{cases} A + e^{-|\Delta t|/\tau_+} & \text{if } \Delta t > 0 \\ A - e^{-|\Delta t|/\tau_-} & \text{if } \Delta t < 0 \end{cases}$$

where A is the scaling factor, τ is the time constant, + means LTP and – means LTD. The net effect of spike train segment pairs was estimated by combining the contributions of all spike pairs multiplicatively²⁸: $1 + \Delta w = \prod_{ij} (1 + \Delta w_{ij})$. Suppression time constants for the pre- and postsynaptic neurons, τ_s^{pre} (34 ms) and τ_s^{post} (75 ms), were determined from the '2/1' and '1/2' triplet experiments, respectively, chosen to minimize mean prediction error: |predicted effect – measured effect|. To predict the effects using the independent model, ϵ^{pre} and ϵ^{post} were set to one regardless of the inter-spike interval.

Alternatively, the contributions of different spike pairs can be combined additively^{11–13,28}: $\Delta w = \sum_{ij} \Delta w_{ij}$. Under the additive model, $\tau_s^{\text{pre}} = 28$ ms, $\tau_s^{\text{post}} = 88$ ms. For quadruplet experiments, the mean prediction errors were $16.1 \pm 3.0\%$ (s.e.m.) (suppression model) and $29.4 \pm 5.7\%$ (independent model). For natural spike train segments, mean prediction errors were $19.5 \pm 2.5\%$ (suppression model) and $45.5 \pm 9.3\%$ (independent model). These errors were not significantly different from those of the multiplicative model (see legends of Figs 3c, 3d and 4c, 4d; $P > 0.25$, t -test).

Received 17 December 2001; accepted 15 January 2002.

- Levy, W. B. & Steward, O. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience* **8**, 791–797 (1983).
- Gustafsson, B., Wigstrom, H., Abraham, W. C. & Huang, Y. Y. Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J. Neurosci.* **7**, 774–780 (1987).
- Markram, H., Lubke, J., Frotscher, M. & Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**, 213–215 (1997).
- Debanne, D., Gähwiler, B. H. & Thompson, S. M. Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J. Physiol. (Lond)* **507**, 237–247 (1998).
- Zhang, L. L., Tao, H. W., Holt, C. E., Harris, W. A. & Poo, M. A critical window for cooperation and competition among developing retinotectal synapses. *Nature* **395**, 37–44 (1998).
- Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
- Egger, V., Feldmeyer, D. & Sakmann, B. Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in rat barrel cortex. *Nature Neurosci.* **2**, 1098–1105 (1999).
- Feldman, D. E. Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron* **27**, 45–56 (2000).
- Nishiyama, M., Hong, K., Mikoshiba, K., Poo, M. M. & Kato, K. Calcium stores regulate the polarity and input specificity of synaptic modification. *Nature* **408**, 584–588 (2000).
- Boettiger, C. A. & Doupe, A. J. Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron* **31**, 809–818 (2001).
- Kempster, R., Gerstner, W. & van Hemmen, J. L. Hebbian learning and spiking neurons. *Phys. Rev. E* **59**, 4498–4514 (1999).
- Roberts, P. D. Computational consequences of temporally asymmetric learning rules: I. Differential hebbian learning. *J. Comput. Neurosci.* **7**, 235–246 (1999).
- Song, S., Miller, K. D. & Abbott, L. F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neurosci.* **3**, 919–926 (2000).
- Somogyi, P., Tamas, G., Lujan, R. & Buhl, E. H. Salient features of synaptic organisation in the cerebral cortex. *Brain Res. Rev.* **26**, 113–135 (1998).
- Tsodyks, M. V. & Markram, H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl Acad. Sci. USA* **94**, 719–723 (1997).
- Varela, J. A. et al. A quantitative description of short-term plasticity at excitatory synapses in layer II/III of rat primary visual cortex. *J. Neurosci.* **17**, 7926–7940 (1997).
- Zucker, R. S. Calcium- and activity-dependent synaptic plasticity. *Curr. Opin. Neurobiol.* **9**, 305–313 (1999).
- Zorumski, C. F. & Thio, L. L. Properties of vertebrate glutamate receptors: calcium mobilization and desensitization. *Prog. Neurobiol.* **39**, 295–336 (1992).
- Stuart, G. J. & Sakmann, B. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature* **367**, 69–72 (1994).
- Malenka, R. C. & Nicoll, R. A. Long-term potentiation—a decade of progress? *Science* **285**, 1870–1874 (1999).
- Lisman, J. E. & Zhabotinsky, A. M. A model of synaptic memory: a CaMKII/PP1 switch that potentiates transmission by organizing an AMPA receptor anchoring assembly. *Neuron* **31**, 191–201 (2001).
- Senn, W., Markram, H. & Tsodyks, M. An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Comput.* **13**, 35–67 (2001).
- Rao, R. P. & Sejnowski, T. J. Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Comput.* **13**, 2221–2237 (2001).
- Yao, H. & Dan, Y. Stimulus timing-dependent plasticity in cortical processing of orientation. *Neuron* **32**, 315–323 (2001).
- Dobrunz, L. E. & Stevens, C. F. Response of hippocampal synapses to natural stimulation patterns. *Neuron* **22**, 157–166 (1999).

26. Connors, B. W. & Gutnick, M. J. Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci.* **13**, 99–104 (1990).
27. Stanley, G. B., Li, F. F. & Dan, Y. Reconstruction of natural scenes from ensemble response in the lateral geniculate nucleus. *J. Neurosci.* **19**, 8036–8042 (1999).
28. van Rossum, M. C., Bi, G. Q. & Turrigiano, G. G. Stable Hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* **20**, 8812–8821 (2000).

Acknowledgements

We thank K. Djupsund for help in recording natural spike trains, and J. Long, M. Frerking and R. S. Zucker for discussions. This work was supported by grants from the National Science Foundation and National Eye Institute. R.C.F. is a recipient of the Howard Hughes Predoctoral Fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.D.
(e-mail: ydan@uclink4.berkeley.edu).

Rac function and regulation during *Drosophila* development

Satoko Hakeda-Suzuki*, Julian Ng†, Julia Tzu†, Georg Dietzl*, Yan Sun*, Matthew Harms†, Tim Nardine†, Liqun Luo†‡ & Barry J. Dickson*

* Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria

† Department of Biological Sciences, and ‡ Neurosciences Program, Stanford University, Stanford, California 94035, USA

Rac GTPases regulate the actin cytoskeleton to control changes in cell shape^{1,2}. To date, the analysis of Rac function during development has relied heavily on the use of dominant mutant isoforms. Here, we use loss-of-function mutations to show that the three *Drosophila* Rac genes, *Rac1*, *Rac2* and *Mtl*, have overlapping functions in the control of epithelial morphogenesis, myoblast fusion, and axon growth and guidance. They are not required for the establishment of planar cell polarity, as had been suggested on the basis of studies using dominant mutant isoforms^{3,4}. The guanine nucleotide exchange factor, Trio, is essential for Rac function in axon growth and guidance, but not for epithelial morphogenesis or myoblast fusion. Different Rac activators thus act in different developmental processes. The specific cellular response to Rac activation may be determined more by the upstream activator than the specific Rac protein involved.

In *Drosophila*, studies using constitutively active and dominant negative mutants have implicated Rac1 in closure of the dorsal epidermis⁵, myoblast fusion⁶, the establishment of planar cell polarity^{3,4}, and the control of axon growth⁶ and guidance⁷. Each of these processes requires dynamic remodelling of the actin cytoskeleton, although the extracellular signals and the cellular responses involved seem to be different in each case. Given the ability of dominant mutant Rac proteins to interfere with cytoskeletal dynamics¹, it is not surprising to find that they perturb each of these processes. But are endogenous Rac proteins actually required for these processes, and if so, which proteins are involved, and how are they regulated? These questions cannot be answered using dominant mutant proteins. They require the phenotypic analysis of loss-of-function mutations in each of the endogenous Rac genes.

The *Drosophila* genome contains two highly similar Rac genes, *Rac1* and *Rac2* (refs 5, 6, 8, 9). A third gene, *Mtl*, encodes a closely

related GTPase that is structurally similar to both Rac and Cdc42 GTPases¹⁰, but functionally (as we now show) behaves like Rac1 and Rac2. We therefore refer to *Rac1*, *Rac2* and *Mtl* collectively as the *Drosophila* Rac genes (Fig. 1a). All three genes are ubiquitously expressed during development^{5,6,8,10}. The isolation of loss-of-function *Rac1* and *Rac2* mutations is described in the accompanying paper¹¹. A loss-of-function mutation in the *Mtl* gene was generated by imprecise excision of a P-element inserted in the first non-coding exon. We recovered a 2,068-base pair (bp) deletion that removes the entire *Mtl* open reading frame, but no part of any other predicted gene (Fig. 1b). Animals homozygous for this deletion, *Mtl*^Δ, as well as both *Rac1* and *Rac2* single mutants¹¹, are viable and fertile. The *Rac2 Mtl* double mutant is also viable and fertile. All other combinations are homozygous lethal.

We used these loss-of-function mutations to assess the contribution of each Rac protein to a set of distinct cell-shape changes that occur during *Drosophila* development. We examined embryos lacking both the maternal and zygotic contributions of one or more Rac gene, and also pupae and adults that were homozygous mutant either entirely or in large clones of cells. For pupae and adults, we used both the strong hypomorph *Rac1*¹¹⁰ and the null allele *Rac1*¹¹¹, together with the null deletion alleles for *Rac2* and *Mtl* (*Rac2*^Δ and *Mtl*^Δ). Analyses in the embryo were restricted to the use of the *Rac1*¹¹⁰ allele, as triple mutant embryos could not be recovered using the null allele *Rac1*¹¹¹. Evidently, Rac proteins also have important but still uncharacterized functions during oogenesis and early embryogenesis.

During *Drosophila* embryogenesis, opposing lateral epidermal sheets move towards one another, meeting and fusing seamlessly at the dorsal midline. This process of dorsal closure resembles ventral enclosure in *Caenorhabditis elegans*¹²; and wound healing in vertebrates¹³. It is believed to be driven, at least in part, by an actomyosin contractile ring that assembles at the leading edge^{14,15}, with lamellipodial and filopodial protrusions facilitating adhesion and alignment as the epidermis is sealed¹⁶. Expression of dominant negative Rac1 in epidermal cells prevents formation of the actomyosin cable and completion of dorsal closure⁵, suggesting that at least one endogenous Rac protein might be involved. We determined that all three Rac proteins contribute to dorsal closure (Fig. 1c–g). Triple mutant Rac embryos fail to complete dorsal closure (Fig. 1c, e). Little or no actin accumulation is seen at the leading epidermal edge, and both lamellipodia and filopodia are lacking (Fig. 1g). The underlying amnioserosa cells appear normal. Weaker and less frequent defects are also seen in *Rac1 Rac2* and *Rac1 Mtl* double mutant embryos (Fig. 1c). All remaining single and double mutant embryos successfully complete dorsal closure (Fig. 1c). Dorsal closure thus relies more on Rac1 than either Rac2 or Mtl, although any one of the three is largely sufficient.

Quite different cell-shape changes occur during cell fusion, a striking example of which is the fusion of myoblasts to form multinucleate muscle fibres¹⁷. The role of the actin cytoskeleton in myoblast fusion remains unclear. Most likely, it is involved in the formation of a vesicular prefusion complex that assembles at the apposed plasma membranes¹⁸. Expression of either dominant negative or dominant active Rac1 in *Drosophila* myoblasts blocks their fusion⁶, but here too the precise roles and contributions of individual Rac genes are unknown. We found that little or no myoblast fusion occurs in either *Rac1 Rac2* double mutant (Fig. 1j) or *Rac1 Rac2 Mtl* triple mutant embryos. In contrast, myoblast fusion appears to be complete in *Rac1* and *Mtl* single and double mutants, whereas only a few isolated myoblasts fail to fuse in *Rac2* single mutants and *Rac2 Mtl* double mutants (Fig. 1i). Myoblast fusion thus requires either Rac1 or Rac2, but not Mtl.

Actin rearrangements also underlie the establishment of planar cell polarity (PCP) within an epithelium¹⁹. In *Drosophila*, PCP has been studied most extensively in the context of eye and wing development. Photoreceptors in the eye are arranged in a trapezoi-

dal fashion, giving each ommatidium a specific chirality and orientation. In the wing, each epithelial cell forms a single distally oriented hair, preceded by the assembly of an actin-based 'pre-hair' at the distal vertex of the cell. The involvement of Rac proteins in PCP has been suggested by the finding that both dominant negative and constitutively active forms of Rac1 disrupt ommatidial orientation in the eye⁴, and that dominant negative Rac1 also induces the formation of multiple hairs per cell in the wing³. To test more critically a requirement for Rac proteins in PCP, we examined clones of cells in the eye and the wing that were triply mutant for null alleles of *Rac1*, *Rac2* and *Mtl*. No PCP defects could be detected within these clones in either tissue (Fig. 1l, m). Presumably, the PCP defects previously reported^{3,4} are due to cross-inhibition or cross-activation of other pathways by the dominant mutant Rac proteins used.

The most complex changes in cell shape that occur during development take place in the nervous system, as differentiating neurons extend axons and dendrites towards their specific target cells. Dominant mutant Rac proteins are crude tools with which to address the complexities and subtleties of neuronal differentiation², and it is perhaps not surprising that their use has sometimes led to conflicting results^{6,7}. To begin to tease apart the diverse functions of Rac GTPases in nervous system development, we examined the embryonic central nervous system (CNS) and peripheral nervous

system (PNS), and the adult visual system, in single, double, and triple mutants for *Rac1*, *Rac2* and *Mtl*.

Embryonic CNS axon pathways were examined using anti-Fasciclin II (FasII) monoclonal antibody 1D4. FasII labels axons in three longitudinal fascicles on each side of the midline, and is thus a sensitive marker to detect any misrouting of longitudinal axons across the midline (Fig. 2a, f). Such midline guidance errors occur in 33% of segments in *Mtl* mutant embryos (Fig. 2b, k), but only at very low frequency (less than 2%) in embryos lacking one or both of *Rac1* and *Rac2* (Fig. 2a, k). Mutations in *Rac1*, and to a lesser extent *Rac2*, enhance the frequency of the midline guidance errors in *Mtl* mutant embryos (to 75% and 42% of segments, respectively; Fig. 2c, k).

Axon guidance defects also occur in the visual system of whole-eye Rac mosaics generated using *eyFLP* (ref. 20). In wild-type adults and control mosaics, photoreceptor axons establish precise topographic connections in the lamina and medulla of the optic lobe (Fig. 3a, f). These projection patterns are largely normal in each of the single mutants, and only mild defects occur in *Rac1 Rac2* and *Rac2 Mtl* double mutants (Fig. 3b, d, g). Projection defects are more pronounced in *Rac1 Mtl* double mutants (Fig. 3c, g), and are severe in the triple mutant (Fig. 3e, g). These defects include local disruptions in topographic mapping, and a frequent misrouting of photoreceptor axons around and beyond the medulla (Fig. 3c, e–

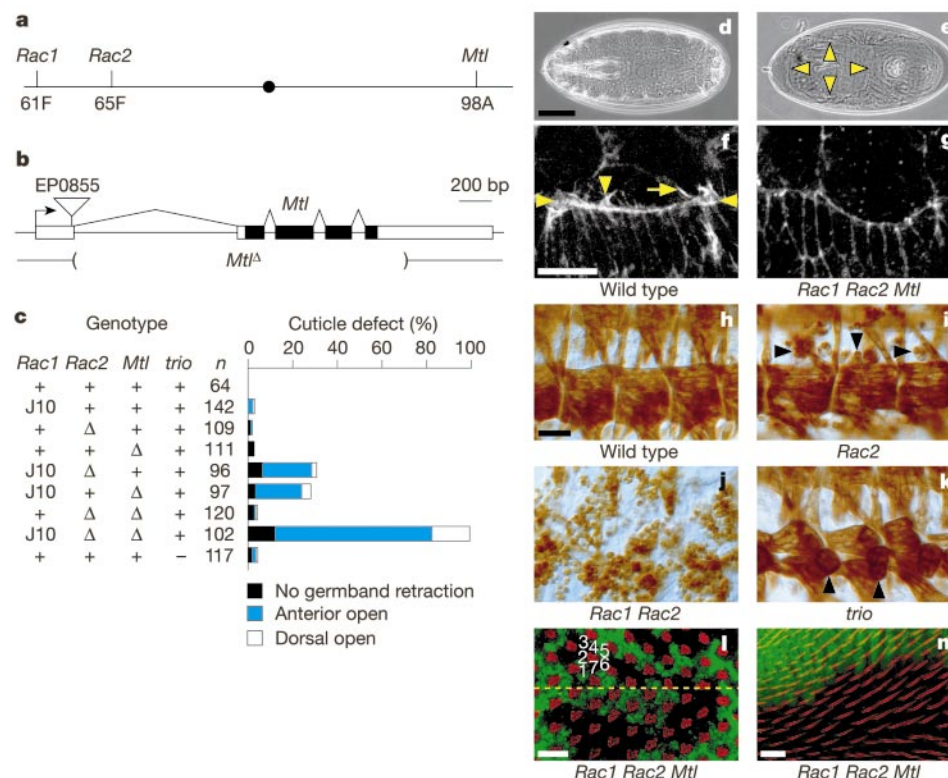


Figure 1 Rac GTPases are required for dorsal closure and myoblast fusion, but not for planar cell polarity. **a**, Location of the three *Drosophila* Rac genes on chromosome 3. **b**, Genomic organization of the *Mtl* locus. Coding regions are shaded. **c**, Quantification of cuticle defects in embryos of the indicated genotype (maternal and zygotic). *trio* embryos were *trio*¹/*trio*¹ maternal and *trio*¹/*trio*⁸ zygotic. **d–g**, Dorsal closure in wild-type (**d**, **f**), and *Rac1*^{J10} *Rac2*^Δ *Mtl*^Δ (**e**, **g**) embryos. **d**, **e**, Cuticle phenotypes of 22-h embryos. The cuticle of *Rac* triple mutant embryos remains open (arrowheads in **e**). **f**, **g**, F-actin accumulation at the leading edge of the epidermis of stage 14 embryos, visualized by rhodamine-phalloidin staining. The actomyosin cable (horizontal arrowheads), lamellipodia (vertical arrowhead) and filopodia (arrow) seen in wild-type embryos (**f**) are mostly lacking in *Rac* triple mutant embryos (**g**). **h–k**, Ventral and lateral musculature of stage 15–16 embryos, stained with anti-muscle myosin monoclonal antibody FMM5.

Arrowheads indicate unfused myoblasts in a *Rac2*^Δ embryo (**f**), and detached myotubes in a *trio* mutant embryo (**k**). **l**, **m**, *Rac1*^{J10} *Rac2*^Δ *Mtl*^Δ triple null clones in the pupal eye (**l**) and wing (**m**), at 75 and 35 h, respectively, after puparium formation, stained with rhodamine-phalloidin (red). Clones were generated using *hsFLP*, and are revealed by the lack of staining for a cytoplasmic β-galactosidase marker (green). **l**, Rhabdomeres of R1–R7 (1–7, red) display the normal trapezoidal arrangement and orientation, with mirror symmetry across the equator (yellow line). Ommatidial polarity is also normal in whole-eye clones generated with *eyFLP* (not shown). Rhabdomere morphology appears normal at this stage, but is often perturbed in adults, as has been observed on expression of dominant negative Rac1 in the eye³⁰. **m**, A single distally oriented pre-hair (red) forms in each cell in the wing disc, both inside and outside the *Rac* triple mutant clones. Scale bars: **d**, 100 μm; **f**, **m**, 10 μm; **h**, **i**, 20 μm.

g). Specification of photoreceptor cell fate appears to be normal, even in triple mutant clones (Fig. 1l and data not shown). The projection defects observed in the triple mutant could be rescued by reintroducing either *Rac1* or *Mtl* specifically in the eye using a *GMR* transgene (Fig. 3g).

Together, these data establish a critical role for endogenous Rac proteins in axon guidance. The three Rac proteins have overlapping functions in axon guidance in both the CNS and visual system, just as they do in the mushroom bodies¹¹. Nevertheless, some degree of specialization can be discerned. For example, axon guidance at the CNS midline depends more on *Mtl* than *Rac1*, whereas *Rac1* is more important than *Mtl* in the mushroom bodies.

Whereas single and double mutant embryos reveal a role for Rac proteins in axon guidance, triple mutant embryos demonstrate the essential function of Rac proteins in axon growth. Severe growth defects occur in *Rac1 Rac2 Mtl* homozygous mutant embryos

(Fig. 2d, i, k). In the CNS, FasII-positive axons rarely extend from one segment into the next (Fig. 2d), and very few sensory axons from the PNS reach the CNS (Fig. 2i). Specification of neuronal and glial cell fate appears relatively normal (not shown), as does dendritic growth and morphology (Fig. 2i). We quantified these axon growth defects by determining the frequency with which axons from the dorsal cluster of sensory neurons reach the lateral cluster on their path towards the CNS (Fig. 2g, k). This analysis of the PNS confirmed the general impression gained from the CNS: axon stalling is severe in *Rac1 Rac2 Mtl* triple mutants, occurs occasionally in *Rac1 Mtl* double mutants, and is rare in all other combinations (Fig. 2h, i, k).

These data demonstrate that axon growth also requires Rac activity, but only at a very low level. This activity can be provided by any one of the three Rac proteins alone, although *Rac2* is less effective than either *Rac1* or *Mtl*. Axon growth is thus maintained at levels of Rac activity that are insufficient for accurate guidance, consistent with the idea that a low level of Rac activity is essential to drive the growth cone forward, while spatially restricted bursts of high activity may be needed to turn it. In the accompanying paper¹¹, we take this model one step further by showing that axon branching requires even higher levels of Rac activity.

Endogenous Rac GTPases thus function in morphogenesis of the epidermis, mesoderm, and nervous system. Are they regulated by the same or different upstream activators in each of these tissues? The guanine nucleotide exchange factor Trio activates *Rac1*, *Rac2* and *Mtl* *in vitro*¹⁰, and loss of *trio* function in the visual system results in projection errors of photoreceptor axons similar to those observed in *Rac* triple mutants¹⁰. Axon guidance errors and occasional stalling defects also occur in embryos lacking zygotic *trio* function^{21–23}. Axon stalling becomes severe in both the CNS (Fig. 2e) and PNS (Fig. 2j, k) if the maternal *trio* function is also eliminated. As with the Rac proteins, low levels of Trio activity are sufficient but essential for axon growth. This critical requirement for Trio in axon growth is particularly striking, given that the *Drosophila* genome encodes at least 22 other Rho family GTPase exchange factors, several of which are also expressed in the developing nervous system (ref. 24 and S.H.-S. and B.J.D., unpublished data).

In the embryonic nervous system and adult visual system, loss of *trio* function thus results in defects remarkably similar to those observed upon loss of *Rac* function, consistent with the idea that Trio and Rac proteins act in a common pathway *in vivo*. We performed an epistasis experiment to test this. Overexpression of the Trio GEF1 domain using the eye-specific *GMR* promoter results in a severely disrupted eye morphology (Fig. 4a) and highly aberrant photoreceptor axon projections (Fig. 4f; see also ref. 10). If Trio signals through Rac proteins *in vivo*, then these defects should be dependent on *Rac* function. This is indeed the case. Both the eye morphology and axon projection defects are almost completely suppressed in animals homozygous for loss-of-function mutations in either *Rac1* or *Rac2* (Fig. 4b, c, g, h). *Mtl* alone does not suppress this *trio* gain-of-function phenotype (Fig. 4d, i). The *Rac1 Rac2 Mtl* triple mutant phenotype is completely epistatic to the *trio* gain-of-function phenotype (Figs 4e, j and 3e). These data demonstrate that Trio GEF1 does indeed act through Rac proteins *in vivo*, and further suggest that *Rac1* and *Rac2* are its preferred substrates. The *trio* loss-of-function phenotype is however much more severe than the *Rac1 Rac2* double mutant phenotype (Figs 3b, g and 2a, e, k), suggesting that endogenous Trio may also activate *Mtl*, at least when *Rac1* and *Rac2* are lacking.

Having identified Trio as the primary activator of Rac proteins during axon growth, we next investigated whether Trio is required for any of the other Rac functions. Dorsal closure occurs normally in embryos lacking both maternal and zygotic *trio* function (Fig. 1c). Myoblast fusion also appears complete in these embryos, but myotubes often fail to attach themselves correctly to the epidermis

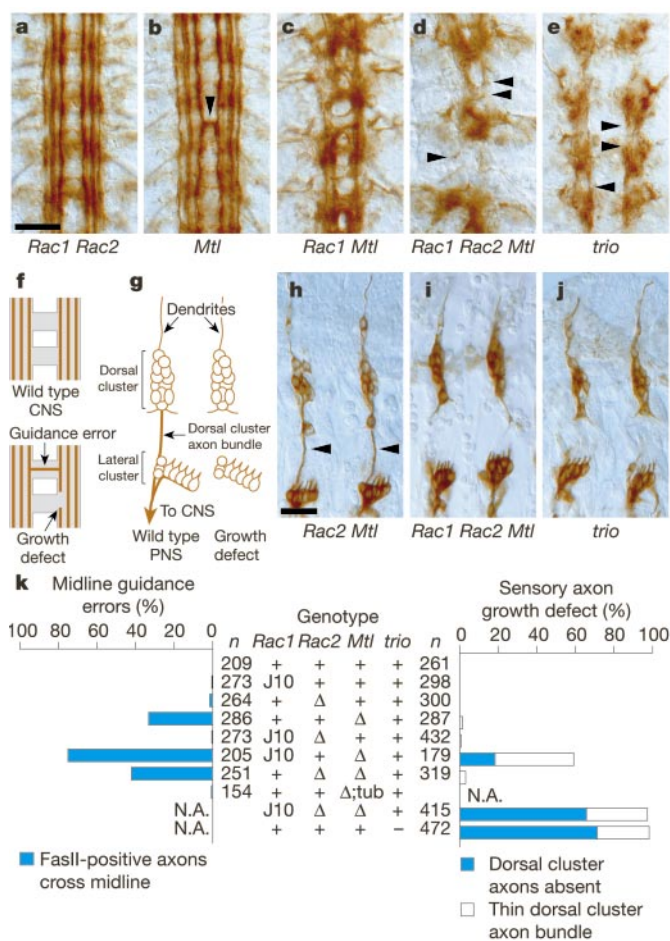


Figure 2 Axon growth and guidance defects in *Rac* mutant embryos. **a–e, h–j**, Late stage 16 embryos stained with anti-FasII monoclonal antibody 1D4 (**a–e**) or 22C10 (**h–j**). Stained with 1D4, the *Rac1^{J10} Rac2^Δ* double mutant (**a**) is identical to wild type. In *Mtl^Δ* (**b**) and *Rac1^{J10} Mtl^Δ* (**c**) mutants, some longitudinal axons are misrouted across the midline (arrowhead in **b**). Severe axon stalling is observed in *Rac1^{J10} Rac2^Δ Mtl^Δ* (**d**) and *trio* mutant (**e**) embryos (arrowheads). Some of the stalled longitudinal axons are directed across the midline in *Rac* triple mutants (**d**), but not in *trio* mutants (**e**). **f**, Schematic of CNS axon pathways, with the three FasII-positive longitudinal axon fascicles highlighted. **g**, Schematic of dorsal and lateral peripheral nervous system (PNS). **h**, A *Rac2^Δ Mtl^Δ* mutant, showing the normal projection of dorsal cluster sensory axons (arrowheads). These axons fail to extend in *Rac1^{J10} Rac2^Δ Mtl^Δ* mutants (**i**) and *trio* mutant (**j**) embryos. Panels **h–j** are photomontages of images acquired in multiple focal planes. **k**, Embryos were scored for midline guidance errors or sensory axon growth defects. *n* indicates the number of segments or hemisegments scored, respectively. *tub*, presence of a *tub-Mtl* transgene; NA, not applicable. Scale bars: 20 μ m.

(Fig. 1k). Thus, although it is expressed in both the epidermis and mesoderm^{21–23}, Trio is not required for either dorsal closure or myoblast fusion.

We have shown here that endogenous Rac proteins control cell-sheet spreading, cell fusion, and axon growth and guidance, and in the accompanying paper¹¹ we show that they also regulate axon branching. Each of these processes involves its own characteristic restructuring of the cytoskeleton, and hence is likely to be mediated by a different set of Rac effectors¹¹. What determines which of these effector pathways will be stimulated when Rac proteins are activated? One possibility would be that distinct Rac proteins have distinct effectors. This may well be the case for myoblast fusion, which can be mediated by Rac1 or Rac2, but not Mtl. However, we

find that in most cases Rac1, Rac2 and Mtl have largely overlapping functions, indicating that they also share a common set of effectors. A similar pattern of overlapping functions in diverse processes has also recently been reported for the three *C. elegans* Rac genes^{25,26}. In general, the cellular response is therefore unlikely to be dictated by the specific Rac protein involved. Our results suggest an alternative possibility. We find that Trio, despite its widespread expression, is required for only a limited subset of Rac functions. This suggests that the set of effectors a Rac protein engages, and hence the cellular response it induces, might also depend on how or where it has been activated. Trio, for example, might activate Rac proteins to a level, for a duration, or in a subcellular location, that allows it to stimulate only those effector pathways that control motility and guidance.

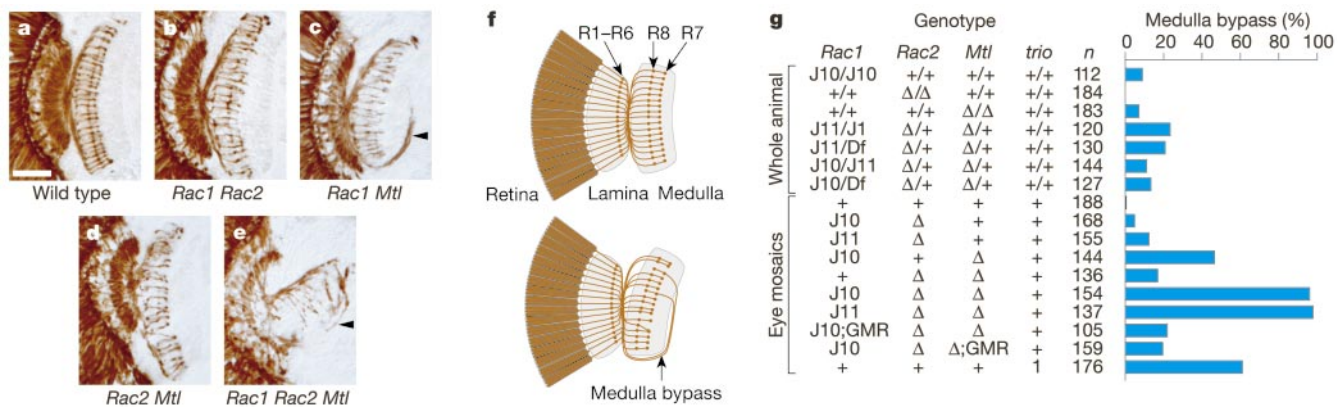


Figure 3 Photoreceptor axon projections in *Rac* mutants. **a–e**, Horizontal sections of adult heads of the indicated genotypes. All animals carried the *glass-lacZ* reporter and were stained with anti-β-galactosidase antibodies to visualize photoreceptors and their axons. Genotypes are as follows: **a**, *eyFLP2; FRT80B/M(3)RpS17^Δ FRT80B*; **b**, *eyFLP2; Rac1^{J11} Rac2^Δ FRT80B/M(3)RpS17^Δ FRT80B*; **c**, *eyFLP2; Rac1^{J10} FRT80B Mtl^Δ/M(3)RpS17^Δ FRT80B Mtl^Δ*; **d**, *eyFLP2; Rac2^Δ FRT80B Mtl^Δ/M(3)RpS17^Δ FRT80B Mtl^Δ*; **e**, *eyFLP2; Rac1^{J11} Rac2^Δ FRT80B Mtl^Δ/M(3)RpS17^Δ FRT80B Mtl^Δ*. No stalling of photoreceptor axons can be detected, but such defects cannot be excluded, in particular for the triple mutant (**e**). Arrowheads in **c** and **e** indicate axons showing the medulla bypass phenotype. These axons are not entirely contained within the sections shown. **f**, Schematic diagram of axonal projections of wild-type (top) and *Rac* or *trio* mutant (bottom) photoreceptors in the adult visual system. The termini of R1–R8 axons are indicated. **g**, Frequency of photoreceptor axons misrouting around the medulla (medulla bypass). Animals were either entirely mutant for the indicated allelic combination (upper lines), or carried whole-eye clones generated using *eyFLP* (ref. 20) (lower lines). Data for *trio* are from ref. 10. Df, *Df(3)Rac1* (ref. 11). GMR indicates the presence of either a *GMR-Rac1* or a *GMR-Mtl* transgene. The *Rac1^{J11}* allele has the same phenotypic strength as the deficiency (lines 4 versus 5, and 6 versus 7). *Rac1^{J10}* appears to be only slightly weaker than both *Rac1^{J11}* (lines 4 versus 6, 9 versus 10, and 13 versus 14) and the deficiency (lines 5 versus 7). These genetic data confirm that *Rac1^{J11}* is a null allele, whereas *Rac1^{J10}* behaves as a strong hypomorph¹¹. Scale bar: 40 μm.

(bottom) photoreceptors in the adult visual system. The termini of R1–R8 axons are indicated. **g**, Frequency of photoreceptor axons misrouting around the medulla (medulla bypass). Animals were either entirely mutant for the indicated allelic combination (upper lines), or carried whole-eye clones generated using *eyFLP* (ref. 20) (lower lines). Data for *trio* are from ref. 10. Df, *Df(3)Rac1* (ref. 11). GMR indicates the presence of either a *GMR-Rac1* or a *GMR-Mtl* transgene. The *Rac1^{J11}* allele has the same phenotypic strength as the deficiency (lines 4 versus 5, and 6 versus 7). *Rac1^{J10}* appears to be only slightly weaker than both *Rac1^{J11}* (lines 4 versus 6, 9 versus 10, and 13 versus 14) and the deficiency (lines 5 versus 7). These genetic data confirm that *Rac1^{J11}* is a null allele, whereas *Rac1^{J10}* behaves as a strong hypomorph¹¹. Scale bar: 40 μm.

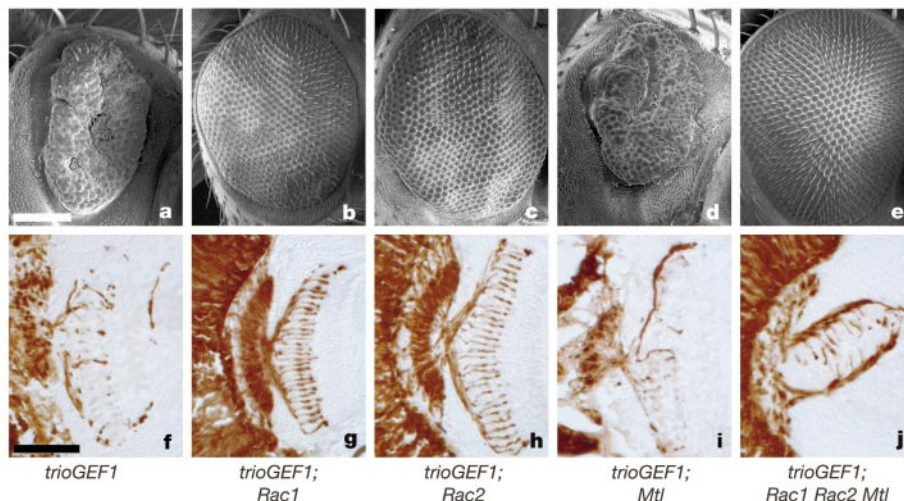


Figure 4 *Rac1* and *Rac2* suppress a *trio* gain-of-function phenotype. **a–j**, Scanning electron micrographs of eyes (**a–e**) and horizontal head sections (**f–j**) of animals of the following genotypes: **a, f**, *GMR-trioGEF1/+*; **b, g**, *GMR-trioGEF1/+; Rac1^{J10}/Rac1^{J10}*; **c, h**, *GMR-trioGEF1/+; Rac2^Δ/Rac2^Δ*; **d, i**, *GMR-trioGEF1/+; Mtl^Δ/Mtl^Δ*; **e, j**, *eyFLP2; GMR-trioGEF1/+; Rac1^{J10} Rac2^Δ FRT80B Mtl^Δ/M(3)RpS17^Δ FRT80B Mtl^Δ*. Scale bars: **a**, 100 μm; **f**, 40 μm.

c, h, *GMR-trioGEF1/+; Rac2^Δ/Rac2^Δ*; **d, i**, *GMR-trioGEF1/+; Mtl^Δ/Mtl^Δ*; **e, j**, *eyFLP2; GMR-trioGEF1/+; Rac1^{J10} Rac2^Δ FRT80B Mtl^Δ/M(3)RpS17^Δ FRT80B Mtl^Δ*. Scale bars: **a**, 100 μm; **f**, 40 μm.

Exploring the basis for specificity in Rac function is an important task for the future. □

Methods

Mosaic analysis

Eye-specific mosaics for chromosome arm 3L were generated using *eyFLP* and *FRT80B* as described²⁰, using *M(3)RpS17⁴* to enhance clone size. For wing and eye clones generated using *hsFLP*, heat shocks of 1 h at 38 °C were administered at 24–48 h and again at 48–72 h of development. For germline clones, *hsFLP* was used together with an *ovo^{D1}* insertion²⁷ on the *FRT80B* chromosome (gift of K. Basler). Heterozygous third instar larvae were heat shocked for 35 min at 39 °C, and adult females crossed to males carrying the appropriate third chromosome over a TM3, *Ubx-lacZ* balancer. In order to generate germline clones in a *Mtl^Δ* homozygous background, *Mtl^Δ* was first recombined onto the *ovo^{D1} FRT80B* chromosome by FLP-induced germline recombination in males.

Histology

Embryos were fixed and stained with monoclonal antibodies 1D4, 22C10, or FMM5, as described²⁸. Embryonic cuticle preparations and F-actin staining were performed as described⁵, using rhodamine-conjugated phalloidin (Molecular Probes, 4 U ml⁻¹). Pupal wings and eyes were prepared and stained with rhodamine-phalloidin and mouse anti-β-galactosidase (Promega, 1:200), as described^{29,30}. Adult head sections were prepared and stained as described²⁰.

Received 1 August 2001; accepted 31 January 2002.

- Hall, A. Rho GTPases and the actin cytoskeleton. *Science* **279**, 509–514 (1998).
- Luo, L. Rho GTPases in neuronal morphogenesis. *Nature Rev. Neurosci.* **1**, 173–180 (2000).
- Eaton, S., Wepf, R. & Simons, K. Roles for Rac1 and Cdc42 in planar polarization and hair outgrowth in the wing of *Drosophila*. *J. Cell Biol.* **135**, 1277–1289 (1996).
- Fanto, M., Weber, U., Strutt, D. I. & Mlodzik, M. Nuclear signaling by Rac and Rho GTPases is required in the establishment of epithelial planar polarity in the *Drosophila* eye. *Curr. Biol.* **10**, 979–988 (2000).
- Harden, N., Loh, H. Y., Chia, W. & Lim, L. A dominant inhibitory version of the small GTP-binding protein Rac disrupts cytoskeletal structures and inhibits developmental cell shape changes in *Drosophila*. *Development* **121**, 903–914 (1995).
- Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
- Kaufmann, N., Wills, Z. P. & Van Vactor, D. *Drosophila* Rac1 controls motor axon guidance. *Development* **125**, 453–461 (1998).
- Hariharan, I. K. et al. Characterization of Rho GTPase family homologues in *Drosophila melanogaster*: overexpressing Rho1 in retinal cells causes a late developmental defect. *EMBO J.* **14**, 292–302 (1995).
- Adams, M. D. et al. The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Newsome, T. P. et al. Trio combines with Dock to regulate Pak activity during photoreceptor axon pathfinding in *Drosophila*. *Cell* **101**, 283–294 (2000).
- Ng, J. et al. Rac GTPases control axon growth, guidance and branching. *Nature* **416**, 442–447 (2002).
- Williams-Masson, E. M., Malik, A. N. & Hardin, J. An actin-mediated two-step mechanism is required for ventral enclosure of the *C. elegans* hypodermis. *Development* **124**, 2889–2901 (1997).
- Jacinto, A., Martinez-Arias, A. & Martin, P. Mechanisms of epithelial fusion and repair. *Nature Cell Biol.* **3**, E117–E123 (2001).
- Young, P. E., Richman, A. M., Ketchum, A. S. & Kiehart, D. P. Morphogenesis in *Drosophila* requires nonmuscle myosin heavy chain function. *Genes Dev.* **7**, 29–41 (1993).
- Edwards, K. A., Demsky, M., Montague, R. A., Weymouth, N. & Kiehart, D. P. GFP-moesin illuminates actin cytoskeleton dynamics in living tissue and demonstrates cell shape changes during morphogenesis in *Drosophila*. *Dev. Biol.* **191**, 103–117 (1997).
- Jacinto, A. et al. Dynamic actin-based epithelial adhesion and cell matching during *Drosophila* dorsal closure. *Curr. Biol.* **10**, 1420–1426 (2000).
- Wakelam, M. J. The fusion of myoblasts. *Biochem. J.* **228**, 1–12 (1985).
- Doberstein, S. K., Fetter, R. D., Mehta, A. Y. & Goodman, C. S. Genetic analysis of myoblast fusion: blown fuse is required for progression beyond the prefusion complex. *J. Cell Biol.* **136**, 1249–1261 (1997).
- Shulman, J. M., Perrimon, N. & Axelrod, J. D. Frizzled signaling and the developmental control of cell polarity. *Trends Genet.* **14**, 452–458 (1998).
- Newsome, T. P., Asling, B. & Dickson, B. J. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851–860 (2000).
- Awasaki, T. et al. The *Drosophila* trio plays an essential role in patterning of axons by regulating their directional extension. *Neuron* **26**, 119–131 (2000).
- Bateman, J., Shu, H. & Van Vactor, D. The guanine nucleotide exchange factor trio mediates axonal development in the *Drosophila* embryo. *Neuron* **26**, 93–106 (2000).
- Liebl, E. C. et al. Dosage-sensitive, reciprocal genetic interactions between the Abl tyrosine kinase and the putative GEF trio reveal trio's role in axon pathfinding. *Neuron* **26**, 107–118 (2000).
- Sone, M. et al. Still life, a protein in synaptic terminals of *Drosophila* homologous to GDP-GTP exchangers. *Science* **275**, 543–547 (1997).
- Reddien, P. W. & Horvitz, H. R. CED-2/CrkII and CED-10/Rac control phagocytosis and cell migration in *Caenorhabditis elegans*. *Nature Cell Biol.* **2**, 131–136 (2000).
- Lundquist, E. A., Reddien, P. W., Hartwig, E., Horvitz, H. R. & Bargmann, C. I. Three *C. elegans* Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. *Development* **128**, 4475–4488 (2001).
- Chou, T. B., Noll, E. & Perrimon, N. Autosomal P[ovoD1] dominant female-sterile insertions in *Drosophila* and their use in generating germ-line chimeras. *Development* **119**, 1359–1369 (1993).
- Rajagopalan, S., Vivanos, V., Nicolas, E. & Dickson, B. J. Selecting a longitudinal pathway: Robo receptors specify the lateral position of axons in the *Drosophila* CNS. *Cell* **103**, 1033–1045 (2000).

- Winter, C. G. et al. *Drosophila* Rho-associated kinase (Drok) links Frizzled-mediated planar cell polarity signaling to the actin cytoskeleton. *Cell* **105**, 81–91 (2001).
- Chang, H. Y. & Ready, D. F. Rescue of photoreceptor degeneration in rhodopsin-null *Drosophila* mutants by activated Rac1. *Science* **290**, 1978–1980 (2000).

Acknowledgements

We thank H. Halbritter for scanning electron microscopy; P. Rørth for generating and providing the EP0855 line; T. Newsome for the initial characterization of *trio* mutant embryos; and J. Knoblich, K. Nasmyth and members of the Dickson laboratory for comments on the manuscript. This research was supported by funding from Boehringer Ingelheim GmbH (B.J.D.), and grants from the National Institutes of Health (L.L.) and the Human Frontiers of Science Program (B.J.D. and L.L.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.J.D. (e-mail: dickson@nt.imp.univie.ac.at).

Rac GTPases control axon growth, guidance and branching

Julian Ng*, Timothy Nardine*, Matthew Harms*, Julia Tzu*, Ann Goldstein†‡, Yan Sun‡, Georg Dietzl‡, Barry J. Dickson‡ & Liqun Luo*†

* Department of Biological Sciences, and † Neurosciences Program, Stanford University, Stanford, California 94305, USA

‡ Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria

§ These authors contributed equally to this work

Growth, guidance and branching of axons are all essential processes for the precise wiring of the nervous system. Rho family GTPases transduce extracellular signals to regulate the actin cytoskeleton¹. In particular, Rac has been implicated in axon growth and guidance^{2–8}. Here we analyse the loss-of-function phenotypes of three Rac GTPases in *Drosophila* mushroom body neurons. We show that progressive loss of combined Rac1, Rac2 and Mtl activity leads first to defects in axon branching, then guidance, and finally growth. Expression of a Rac1 effector domain mutant that does not bind Pak rescues growth, partially rescues guidance, but does not rescue branching defects of *Rac* mutant neurons. Mosaic analysis reveals both cell autonomous and non-autonomous functions for Rac GTPases, the latter manifesting itself as a strong community effect in axon guidance and branching. These results demonstrate the central role of Rac GTPases in multiple aspects of axon development *in vivo*, and suggest that axon growth, guidance and branching could be controlled by differential activation of Rac signalling pathways.

The *Drosophila* genome has two *Rac* genes that share 92% amino acid sequence identity and have overlapping expression patterns^{2,9,10}. A highly related *Mig-2-like* (*Mtl*) gene, the orthologue of *Caenorhabditis elegans* *mig-2* (ref. 4), is present on the same chromosome^{11,12}. To isolate loss-of-function mutants of *Rac1* and *Rac2*, we generated small deficiencies by means of imprecise excision of nearby P-elements (Fig. 1a, b). The *Rac2^Δ* excision disrupted only the *Rac2* open reading frame (ORF) (Fig. 1a), and hence is a *Rac2*-specific null mutation, but is homozygous viable. The *Df(3)Rac1* excision disrupted the *Rac1* ORF and two adjacent genes (Fig. 1b). *Rac1* point mutations were then recovered from

mutant brains, neuroblast clones, or single neuron clones, are caused by misguidance of MB axons.

Axon growth is least sensitive to loss of Rac GTPases. Defects in axon growth were found mainly in MB neurons homozygous for *Rac1¹¹¹ Rac2^Δ Mtl^Δ* (Fig. 2d). Ball-like axon accumulations were only rarely observed in this genotype. Instead, these axons terminated prematurely or failed to enter FasII-positive regions altogether (Fig. 3n–o). Most of the remaining axons exhibited severe guidance defects, extending their axons in a non-stereotypical fashion. To confirm our interpretation of a growth defect, we examined single-cell clones homozygous for *Rac1¹¹¹ Rac2^Δ Mtl^Δ* in brain hemispheres in which most of the FasII-positive γ -axons (representing non-clonal tissue) were correctly patterned as in wild type. We found that 55% of homozygous *Rac1¹¹¹ Rac2^Δ Mtl^Δ* single-cell clones exhibited axon-stalling defects, mostly at the peduncle (Figs 3p, q and 4b), as compared with less than 5% in *Rac1¹¹¹ Rac2^Δ* single-cell clones ($n > 100$). These data indicate that Rac1, Rac2

and Mtl collaborate to control axon growth in a cell-autonomous manner.

The dendritic region of MB neuroblast clones homozygous for *Rac1¹¹¹ Rac2^Δ Mtl^Δ* still possessed mCD8–GFP-positive neurites, which were presumably contributed by the initial neurite outgrowth from the cell body and elaboration of MB dendrites (Fig. 3m–o). Quantitative analysis of single-cell *Rac1¹¹¹ Rac2^Δ Mtl^Δ* clones revealed significant reduction in both total dendritic length (wild type, $66.3 \pm 5.8 \mu\text{m}$; mutant, $49.6 \pm 4.7 \mu\text{m}$, $n = 16$ and 13, respectively) and number of dendritic segments per neuron (wild type, 18.1 ± 1.8 ; mutant, 11.4 ± 1.6 , $n = 16$ and 13, respectively), indicating that Rac GTPases are also required for dendritic growth and branching.

The differential sensitivity of axon growth, guidance and branching to loss of Rac function could in principle reflect the fact that axons must grow in order to be assayed for guidance and branching defects, and may need correct guidance to reach appropriate branching points. However, such a simple hierarchical model cannot explain our data. If Rac were equally required for growth, guidance and branching, and these processes were simply epistatic to one another as proposed by this hierarchical model, then growth defects should always be more frequent than guidance defects, which in turn should always be more frequent than branching defects. For example, if growth, guidance and branching were equally reduced by 20%, then in a population of 100 axons, 20 would show a growth defect, 16 would show a guidance defect (20% of the 80 axons that grow), and 13 would show a branching defect (20% of the 64 axons that grow and navigate correctly). However, this is not what we observed; instead, we saw a shift from branching to guidance to growth defects as the combined level of wild type Rac is progressively reduced (Fig. 2c–d). That axon branching can be selectively perturbed is best exemplified in genotypes involving a hypomorphic allele *Rac1¹¹⁰* (Fig. 2c, d). Furthermore, axon branching defects could be disrupted without any growth or guidance defects (Fig. 3f–i), whereas guidance could be disrupted without obvious growth defects (Fig. 3l, m). These data strongly suggest that axon growth, guidance and branching are separable events requiring increasing amounts of combined Rac GTPase activity *in vivo*.

One model that could account for these observations is that growth, guidance and branching use different Rac effector pathways. To test this idea, we made use of Rac effector domain mutants. The Rac Phe37Ala mutation (RacF37A) abolishes the ability of activated Rac to induce lamellipodia formation without affecting the Pak/JNK pathway. However, the Rac Tyr40Cys mutation (RacY40C) blocks Rac binding to effectors containing the 'CRIB' motif, including Pak^{19–22}, but does not affect lamellipodia formation^{19,20} (Fig. 4a). Overexpression of analogous Rac1F37A and Rac1Y40C mutants, or wild-type Rac1, did not disrupt MB axon patterning in a wild-type background (Fig. 4c). We could therefore use the MARCM system to express these effector mutants specifically in Rac mutant clones to determine which effector pathways are required for MB axon growth, guidance and branching.

MB axon growth defects in single-cell *Rac1¹¹¹ Rac2^Δ Mtl^Δ* clones were largely rescued by transgenic expression of wild-type Rac1 or Rac1Y40C, but not Rac1F37A (Fig. 4b), indicating that direct binding of CRIB effector proteins are not required for Rac function in axon growth. To assess the effector pathways involved in guidance and branching, we expressed these same transgenes in *Rac1¹¹¹ Rac2^Δ* neuroblast clones. In this background, 81% of axons show mutant phenotypes, predominantly guidance (55%) and branching (24%) defects. Expression of wild-type Rac1 markedly rescued these defects, reducing the fraction of abnormal axon phenotypes to 53% (Fig. 4c). The remaining branching and guidance defects were probably caused by influences of nearby non-clonal MB neurons (see below) heterozygous for *Rac1¹¹¹ Rac2^Δ*, which exhibited a similar degree of branching and guidance defects (Fig. 2c). Neither Rac1Y40C nor Rac1F37A expression reduced the percentage of total

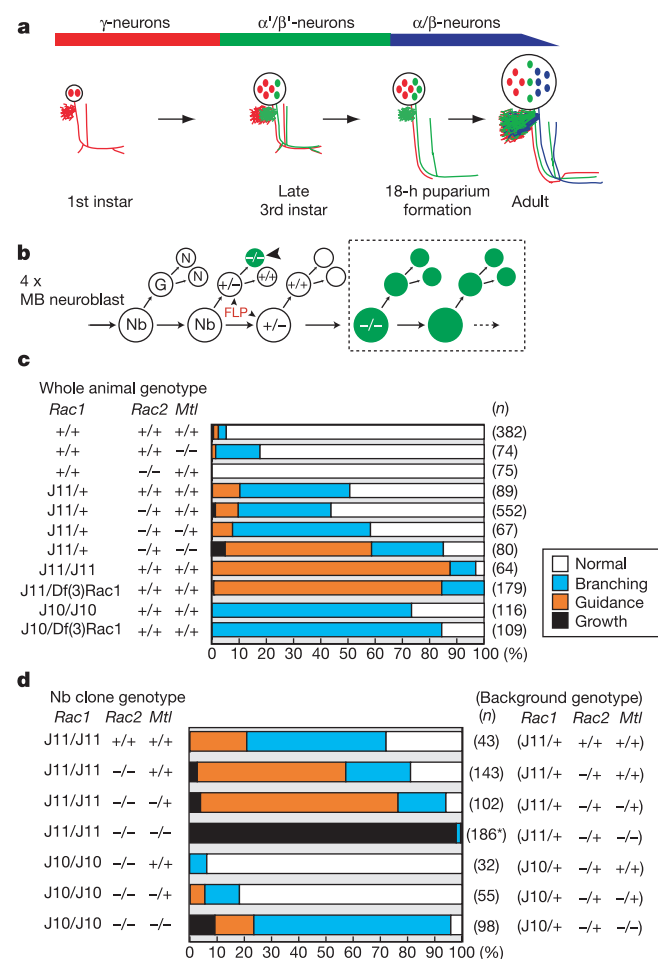


Figure 2 Quantitative phenotypic analysis of Rac GTPases in MB axon branching, guidance and growth. **a**, Schematic summary of MB development¹⁶. **b**, Schematic of MB neuroblast division pattern, illustrating FLP-induced mitotic recombination and the generation of homozygous mutant (–/–) neuroblast clones (boxed) or single-cell MB clones (arrowhead) that are uniquely marked (green) via the MARCM system¹⁸. Nb, neuroblast; G, ganglion mother cell; N, post-mitotic neuron. **c**, Percentages of MB axon branching, guidance and growth defects in adult viable and isogenic *Rac* mutants. **d**, Percentages of branching, guidance and growth defects in *Rac* mutant MB neuroblast clones, generated and labelled via MARCM. Most growth defects in triple mutant neuroblast clones are also associated with severe guidance errors (indicated with an asterisk). n , number of brain hemispheres (**c**) or neuroblast clones (**d**) examined. All chromosomes except *Rac2^Δ* and *Mtl^Δ* are in the *FRT^{2A}* background, which when homozygous consistently gives a small percentage of errors in MB axon development.

axonal defects (78% and 88%, respectively). However, expression of Rac1Y40C (but not Rac1F37A) resulted in a marked shift in the distribution of axonal defects, with most showing branching (45%) rather than guidance (31%) defects. Thus, compared with wild-type Rac1, expression of Rac1Y40C in a Rac mutant background was able to rescue growth, partially rescue guidance, but was unable to rescue branching defects (Fig. 4d).

These results suggest that different downstream effector pathways

mediate axon growth, guidance and branching. In particular, Rac binding of CRIB-domain effectors such as Pak is not required for axon growth, but may contribute to axon guidance and branching. This is consistent with genetic analyses indicating a requirement for *Drosophila* Pak in axon guidance but not growth^{11,23}.

Our mosaic analysis revealed an unexpected degree of cell non-autonomous effects in axon guidance and branching caused by defective Rac activity. If every MB axon were to choose its pathway

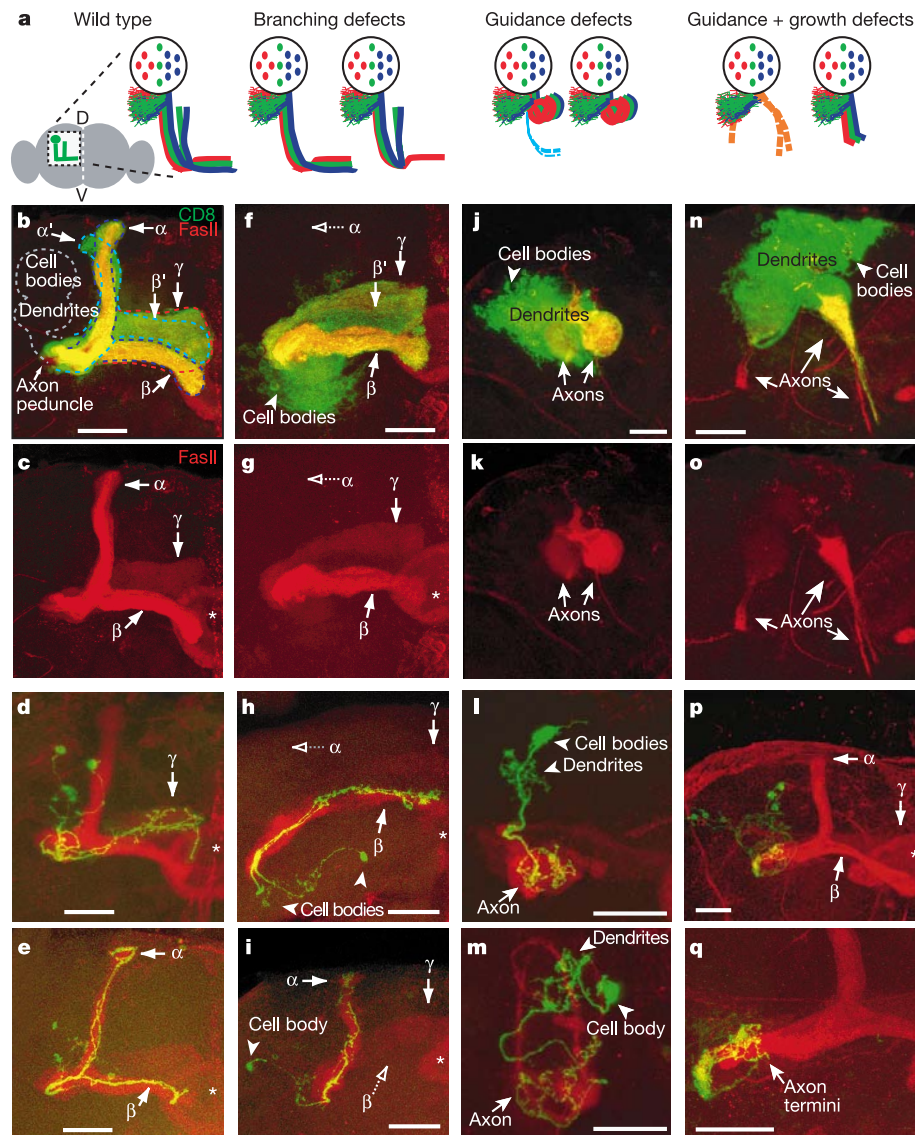


Figure 3 Images of wild-type and *Rac* mutant MB axons. **a**, Schematic summary of branching, guidance and growth defects in *Rac* mutant MB neurons. All drawings and images (**b–q**, Z-projections of confocal sections) are of adult MB of the left hemisphere, roughly from the boxed region in the scheme of the brain on the left. D, dorsal; V, ventral. The white dashed line indicates the midline. Green, neuroblast or single-cell MARCM clones (sometimes multiple single-cell clones) immunostained with anti-CD8 antibodies; red, FasII staining of all MB γ - and α/β -axons. **b–e**, Wild-type MB axons in neuroblast clones (**b**), two single-cell clones of γ -neurons (**d**) or α/β -neurons (**e**). The extent of each axonal lobe is outlined in **b** and labelled with arrows. **f–i**, *Rac* mutant MB neuroblast clones (**f**) and two (**h**) or one (**i**) single-cell clones of α/β -neurons exhibiting axon branching defects. Filled or open arrows point to existing or missing lobes, respectively. **j–m**, *Rac* mutant MB neuroblast clones (**j**) and two-cell (**l**) or single-cell (**m**) clones of γ -neurons exhibiting axon guidance defects. **n, o**, *Rac* mutant MB neuroblast clones exhibiting axon growth and guidance defects. **p, q**, Six single-cell clones exhibiting axon

growth defects in a brain hemisphere where global MB axon patterning is relatively normal, as revealed by FasII staining. **q**, Higher magnification of **p**. Arrow points to axon termini at the peduncle before the branching point. **c, g, k, o** are the same as **b, f, j** and **n**, respectively, showing only FasII staining. Asterisk indicates the ellipsoid body, a central brain structure also stained with FasII. Scale bars, 20 μ m. Genotypes within the clones: **b–e**, *FRT^{2A}*; **f–i**, *Rac1^{J11} Rac2 Δ FRT^{2A}*; **m–q**, *Rac1^{J11} Rac2 Δ FRT^{2A} Mtd^{\Delta}*. Genotypes outside the clones: **b–e**, *FRT^{2A}*; **f–i**, *Rac1^{J11} +, Rac2 Δ /+, FRT^{2A}*; **m–q**, *Rac1^{J11} +, Rac2 Δ /+, FRT^{2A}, Mtd^{\Delta}*. Cell body positions vary as labelled clones can be derived from any of the four MB neuroblasts. In some cases, MB cell bodies, dendrites and axonal peduncles overlap with the axonal lobes in the x–y plane, and were removed from the projection, or replaced with a cartoon (**a**). We did not observe obvious reduction of the number of MB neurons in *Rac1^{J11} Rac2 Δ* neuroblast clones, but observed a slight reduction in *Rac1^{J11} Rac2 Δ Mtd^{\Delta}* neuroblast clones.

independently, then in a brain hemisphere containing one homozygous *Rac1^{J11} Rac2^Δ* neuroblast and three heterozygous neuroblasts, one would expect to observe a mixture of wild-type and mutant trajectories. Remarkably, all FasII-positive axons in the same hemisphere as the mutant neuroblast clone invariably exhibited the same guidance defect (compare Fig. 3k with j). This cannot be explained simply by the guidance defect caused by *Rac1^{J11} Rac2^Δ* heterozygous neurons, as only 8% of such hemispheres exhibited this defect (Fig. 2c). Such a cell non-autonomous effect was also observed in axon branching. If every branching-defective α/β -neuron were to make an independent decision to form either a single dorsal or medial branch, then in neuroblast clones where hundreds of axons are examined together, one would expect to see a thinning of both axonal lobes rather than the absence of a single lobe. Instead, in *Rac1^{J11} Rac2^Δ* neuroblast clones that exhibited branching defects, all MB axons in the neuroblast clone projected either dorsally (one-third) or medially (two-thirds). Most non-clonal axons (as revealed by FasII staining) would always make the same choice as the mutant clone (Fig. 3f, g). Whereas homozygous

mutant axons could induce their heterozygous neighbours to make the same errors, it is possible that heterozygous axons could reduce the error rate of nearby homozygous mutant axons. This could be one explanation for why the phenotypes of homozygous mutant animals are of higher penetrance than those of homozygous mutant neuroblast clones (compare Fig. 2c with d).

These observations suggest a marked community effect in MB axon guidance and branching: axons of mixed genotypes make their choices together. In any given animal, the collective choice of a normal versus mutant projection is likely to be influenced by the severity of the genotype, and the relative number of homozygous versus heterozygous axons. Such collective decision making is probably a result of tight fasciculation among MB axons, which is not disrupted in *Rac* mutants. It will be interesting to test whether this community effect reflects a general feature of axon development in a complex central nervous system environment.

Our analysis of Rac GTPases in MB axon development suggests a mechanistic link between axon growth, guidance and branching. Although there is evidence that axon growth and guidance have different cytoskeletal requirements^{24–26}, their connections are not well understood⁶. Little is known about intracellular signalling mechanisms that regulate axon branching^{27,28}. Here we show that axon growth, guidance and branching require increasing amounts of combined Rac GTPase activity *in vivo* (Figs 2c, d and 4b, d). In the accompanying paper, we report similar differential requirements for Rac activity in embryonic axon growth and guidance¹². The requirement of Rac GTPases for axon outgrowth and guidance in *C. elegans* has also been reported recently⁸. We propose (Fig. 4d) that axon branching, guidance and growth specified by extracellular cues require different amount of Rac GTPase activation in the growth cone, which in turn engage different downstream pathways to specify distinct cytoskeletal changes. □

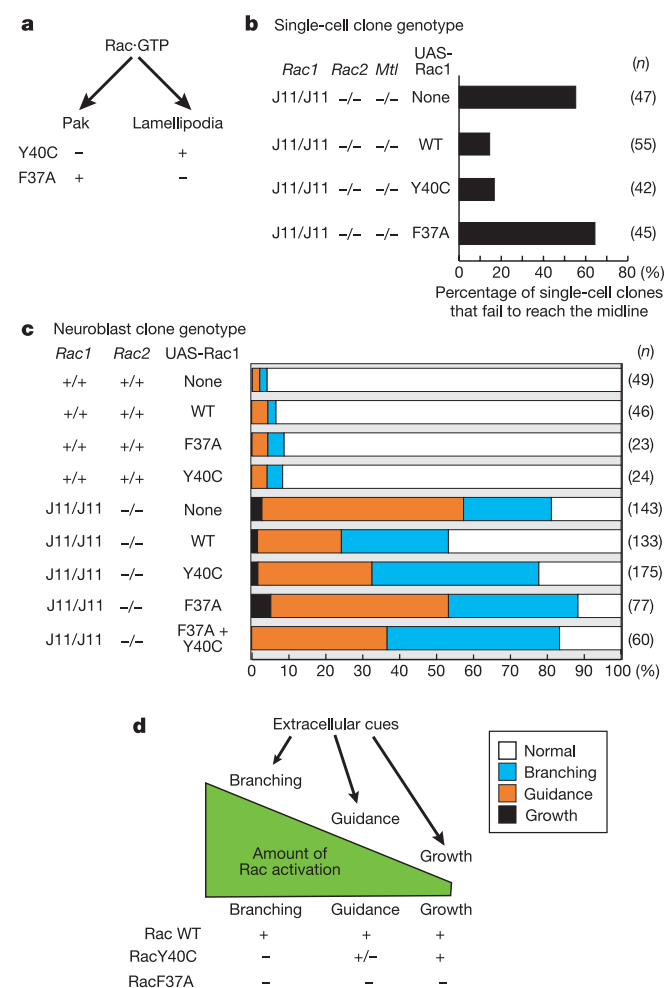


Figure 4 Effector domain mutant analysis. **a**, Specificity of Rac effector domain mutants in activating Pak kinase or inducing lamellipodial formation. **b**, Quantification of axon growth defects of *Rac1^{J11} Rac2^Δ Mtl^Δ* single-cell clones in the absence or presence of *Rac1* transgenes overexpressing wild-type (WT) or effector domain mutants. *n*, number of single-cell γ -neuron clones examined. **c**, Percentage of branching, guidance and growth defects in *Rac1^{J11} Rac2^Δ* neuroblast clones in the absence or presence of *Rac1* transgenes overexpressing wild-type or effector domain mutants. *n*, number of neuroblast clones examined. Also shown are the effects of expressing these transgenes in control flies (*FRT^{2A}*). **d**, Schematic summary and a working hypothesis. See text for details.

Methods

Genetics and molecular biology

P-elements near *Rac1* and *Rac2* were obtained from the Berkeley Drosophila Genome Project. Insertion sites were determined by inverse polymerase chain reaction (PCR). We generated *Rac2* null alleles via imprecise excision of P-element rk519. Imprecise excision of P-element M027 yielded flies carrying a 6.4-kilobase (kb) deletion (which we named *Df(3)Rac1*) from the insertion site to codon 21 of *Rac1*, abolishing *Rac1* function but also removing two other genes (Fig. 1b). *Rac1*-specific mutations were isolated by screening through 8,500 chromosomes mutagenized with EMS for those that failed to complement *Df(3)Rac1* in a *Rac2^Δ* homozygous background. In a *Rac2^Δ* homozygous background, these lethal mutations could be rescued to full viability and fertility with a transgene carrying a genomic fragment of the *Rac1* gene (*Grac1*, Fig. 1b). To generate *Grac1*, a 3.6-kb genomic fragment containing the entire transcription unit and flanking DNA of *Rac1* was subcloned from a P1 clone (DS06962) into the transformation vector pW8 to obtain germline transformants. To determine the molecular identity of *Rac1* mutations, genomic DNA encompassing the *Rac1* ORF was amplified from mutants by PCR, and sequenced. Effector domain mutants were constructed via site-directed mutagenesis, verified by sequencing, subcloned into a vector containing two copies of an amino-terminal Myc tag, and finally into pUAST for germline transformation.

GTP-binding assay

Wild-type *Rac1* and *Rac1^{J11}* were amplified by PCR and subcloned into pGEX 4T-1 to express as glutathione S-transferase (GST) fusion proteins in strain BL21. We performed GTP binding assays according to methods described previously²⁹. [α -³²P]GTP bound to GST beads was used as the baseline of GTP binding. GTP binding of *Rac1* was dependent on Mg^{2+} (data not shown).

MARCM analysis

The MARCM system uses FLP recombinase to induce site-specific mitotic recombination. (FLP, flippase; FRT, FLP recognition target.) Neuroblast and single-cell clones were generated using *hs-FLP* and the 3L *FRT^{2A}*, and visualized by mCD8-GFP expression driven by Gal4-OK107, which labels all MB neurons, as described previously¹⁶. All MARCM clones except those shown in Fig. 3h, i were generated by heat-shock-induced FLP expression in newly hatched larvae, when neuroblast proliferation and clone generation is largely limited to MB neuroblasts^{16,18,30}. Neuroblast clones generated in this fashion comprise approximately one-quarter of all MB neurons in a given brain hemisphere with all three classes of MB neurons. Single-cell clones are exclusively γ -neurons. These are particularly useful for phenotypic analysis in adults, as the larval-specific axon branches are pruned and adult-specific branches are not generated until 4–5 days after clone generation¹⁶, thus minimizing the continuing presence of Rac protein from the time of clone generation to the time of axon re-extension. To generate single-cell

α/β -neuron clones, for studying branching, heat shock was applied at the pupal stage, when MB neuroblasts are again the predominant dividing neuroblast^{18,30}. We also used the MARCM system for Gal4-induced UAS-transgene expression specifically in MB clones for rescue experiments (Fig. 4b, c). Myc-tagged Rac1 or effector domain mutants (Myc-Rac1Y40C or Myc-Rac1F37A) were expressed at similar levels, as judged by Myc-staining.

Immunohistochemistry and image analyses

Isogenic or mosaic mutant brains were dissected within 3–7 days after eclosion, fixed, immunostained for mCD8 and FasII, and imaged as described previously¹⁶. For quantifying MB dendrites, confocal images taken at 0.5 μ m were traced using the NeuroLucida software. The total numbers of branch segments and branch length were quantified using Neuroexplorer (mean $\pm$ standard errors are presented).

Received 1 August 2001; accepted 31 January 2002.

- Hall, A. Rho GTPases and the actin cytoskeleton. *Science* **279**, 509–514 (1998).
- Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).
- Kozma, R., Sarner, S., Ahmed, S. & Lim, L. Rho family GTPases and neuronal growth cone remodelling: relationships between increased complexity induced by Cdc42Hs, Rac1, and acetylcholine and collapse induced by RhoA and lysophosphatidic acid. *Mol. Cell. Biol.* **17**, 1201–1211 (1997).
- Zipkin, I. D., Kindt, R. M. & Kenyon, C. J. Role of a new Rho family member in cell migration and axon guidance in *C. elegans*. *Cell* **90**, 883–894 (1997).
- Kaufmann, N., Wills, Z. P. & Van Vactor, D. *Drosophila* Rac1 controls motor axon guidance. *Development* **125**, 453–461 (1998).
- Luo, L. Rho GTPases in neuronal morphogenesis. *Nature Rev. Neurosci.* **1**, 173–180 (2000).
- Dickson, B. J. Rho GTPases in growth cone guidance. *Curr. Opin. Neurobiol.* **11**, 103–110 (2001).
- Lundquist, E. A., Reddien, P. W., Hartwig, E., Horvitz, H. R. & Bargmann, C. I. Three *C. elegans* Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. *Development* **128**, 4475–4488 (2001).
- Hariharan, I. K. *et al.* Characterization of rho GTPase family homologues in *Drosophila melanogaster*: overexpressing Rho1 in retinal cells causes a late developmental defect. *EMBO J.* **14**, 292–302 (1995).
- Harden, N., Loh, H. Y., Chia, W. & Lim, L. A dominant inhibitory version of the small GTP-binding protein Rac disrupts cytoskeletal structures and inhibits developmental cell shape changes in *Drosophila*. *Development* **121**, 903–914 (1995).
- Newsome, T. *et al.* Trio combines with dock to regulate Pak activity during photoreceptor axon pathfinding in *Drosophila*. *Cell* **101**, 283–294 (2000).
- Hakeda-Suzuki, S. *et al.* Rac function and regulation during *Drosophila* development. *Nature* **416**, 438–442 (2002).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**, 117–127 (1991).
- Worthylake, D. K., Rossman, K. L. & Sondek, J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* **408**, 682–688 (2000).
- Mosteller, R. D., Han, J. & Broek, D. Identification of residues of the H-ras protein critical for functional interaction with guanine nucleotide exchange factors. *Mol. Cell. Biol.* **14**, 1104–1112 (1994).
- Lee, T., Lee, A. & Luo, L. Development of the *Drosophila* mushroom bodies: sequential generation of three distinct types of neurons from a neuroblast. *Development* **126**, 4065–4076 (1999).
- Ito, K., Awano, W., Suzuki, K., Hiromi, Y. & Yamamoto, D. The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurons and glial cells. *Development* **124**, 761–771 (1997).
- Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
- Lamarque, N. *et al.* Rac and Cdc42 induce actin polymerization and G1 cell cycle progression independently of p65PAK and the JNK/SAPK MAP kinase cascade. *Cell* **87**, 519–529 (1996).
- Joneson, T., McDonough, M., Bar-Sagi, D. & Van Aelst, L. RAC regulation of actin polymerization and proliferation by a pathway distinct from Jun kinase. *Science* **274**, 1374–1376 (1996).
- Morreale, A. *et al.* Structure of Cdc42 bound to the GTPase binding domain of PAK. *Nature Struct. Biol.* **7**, 384–388 (2000).
- Owen, D., Mott, H. R., Laue, E. D. & Lowe, P. N. Residues in Cdc42 that specify binding to individual CRIB effector proteins. *Biochemistry* **39**, 1243–1250 (2000).
- Hing, H., Xiao, J., Harden, N., Lim, L. & Zipursky, S. Pak functions downstream of Dock to regulate photoreceptor axon guidance in *Drosophila*. *Cell* **97**, 853–863 (1999).
- Yamada, K. M., Spooner, B. S. & Wessells, N. K. Axon growth: roles of microfilaments and microtubules. *Proc. Natl Acad. Sci. USA* **66**, 1206–1212 (1970).
- Marsh, L. & Letourneau, P. C. Growth of neurites without filopodial or lamellipodial activity in the presence of cytochalasin B. *J. Cell. Biol.* **99**, 2041–2047 (1984).
- Bentley, D. & Toroian-Raymond, A. Disoriented pathfinding by pioneer neurone growth cones deprived of filopodia by cytochalasin treatment. *Nature* **323**, 712–715 (1986).
- Scott, E. K. & Luo, L. How do dendrites take their shape? *Nature Neurosci.* **4**, 359–365 (2001).
- Wang, K. H. *et al.* Biochemical purification of a mammalian slit protein as a positive regulator of sensory axon elongation and branching. *Cell* **96**, 771–784 (1999).
- Knaus, U. G., Heyworth, P. G., Kinsella, B. T., Curnutte, J. T. & Bokoch, G. M. Purification and characterization of Rac2. *J. Biol. Chem.* **267**, 23575–23582 (1992).
- Ito, K. & Hotta, Y. Proliferation pattern of postembryonic neuroblasts in the brain of *Drosophila melanogaster*. *Dev. Biol.* **149**, 134–148 (1992).

Acknowledgements

We thank the Berkeley *Drosophila* Genome Project for providing the P-elements, C. Goodman for the FasII antibody, and members of the Luo laboratory and M. Tessier-Lavigne for critical reading of the manuscript. J. N. was partly supported by a postdoctoral

fellowship from the American Cancer Society California Division. This research was supported by grants from the National Institutes of Health to L.L., and from the Human Frontiers Science Program to L.L. and B.J.D.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.L. (e-mail: lluo@stanford.edu).

Calmodulin interacts with MLO protein to regulate defence against mildew in barley

Min C. Kim*†§, Ralph Panstruga†§, Candace Elliott‡, Judith Müller†, Alessandra Devoto‡, Hae W. Yoon*, Hyeong C. Park*, Moo J. Cho* & Paul Schulze-Lefert†

* Division of Applied Life Science (BK21 program), Plant Molecular Biology & Biotechnology Research Centre, Gyeongsang National University, Chinju 660-701, Korea

† Max-Planck-Institut für Züchtungsforschung, Department of Plant Microbe Interactions, Carl-von-Linné-Weg 10, D-50829 Köln, Germany

‡ The Sainsbury Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK

§ These authors contributed equally to this work

In plants, defence against specific isolates of a pathogen can be triggered by the presence of a corresponding race-specific resistance gene¹, whereas resistance of a more broad-spectrum nature can result from recessive, presumably loss-of-regulatory-function, mutations². An example of the latter are *mlo* mutations in barley, which have been successful in agriculture for the control of powdery mildew fungus (*Blumeria graminis* f. sp. *hordei*; Bgh)³. MLO protein resides in the plasma membrane, has seven transmembrane domains, and is the prototype of a sequence-diversified family unique to plants^{4,5}, reminiscent of the seven-transmembrane receptors in fungi and animals⁵. In animals, these are known as G-protein-coupled receptors and exist in three main families, lacking sequence similarity, that are thought to be an example of molecular convergence⁶. MLO seems to function independently of heterotrimeric G proteins. We have identified a domain in MLO that mediates a Ca²⁺-dependent interaction with calmodulin *in vitro*. Loss of calmodulin binding halves the ability of MLO to negatively regulate defence against powdery mildew *in vivo*. We propose a sensor role for MLO in the modulation of defence reactions.

We adopted a genetic approach to test for the possible involvement of heterotrimeric G proteins in MLO-dependent defence modulation. Unlike metazoans, plants generally have only one gene coding for a canonical heterotrimeric G-protein α -subunit (ref. 7, and J. Glazebrook, personal communication). Southern blot analysis revealed that barley also contains a single-copy gene for G α (*HvG α* ; data not shown). We isolated a full-length *HvG α* complementary DNA and generated three nominally constitutive active variants, *HvG α G48V*, *HvG α R191C* and *HvG α Q223L*, for testing in barley-leaf epidermal cells, using a system for delivering the DNA constructs on microprojectiles⁸. Each of the substitutions involves a residue within the GTP-binding domain that is invariant among G α proteins (Supplementary Information 1), and has been shown to inhibit the intrinsic GTPase activity of animal, fungal and plant G α

proteins^{9–12}. Two other substitutions of single conserved residues were made, HvGα S53C and HvGα R227S, which have each been shown to impair the release of Gβγ from the Gα subunit in animal or yeast Gα proteins, thereby inactivating signalling by way of Gα or Gβγ (Supplementary Information 1)^{13,14}. Plasmids containing the wild-type *HvGα* gene or *HvGα* variants driven by the strong constitutive maize ubiquitin promoter were bombarded together with a *GUS* reporter plasmid used for marking the individual transformed epidermal cells. At 48 h after inoculation with powdery mildew spores, we counted the frequency with which *Bgh* sporelings penetrated the leaf epidermal cell walls, because this is the stage of fungal development that the *mlo* resistance blocks. Neither the wild-type *HvGα* construct nor any of the *HvGα* variants resulted in altered penetration frequencies in the transformed cells, in leaves of either the *Mlo* or *mlo* genetic background (expressed as relative susceptibility in Fig. 1a, b).

We also generated inverted-repeat constructs, *HvGα* double-stranded RNA interference (dsRNAi) and *Mlo* dsRNAi, for hairpin-RNA-mediated silencing of the endogenous *HvGα* and *Mlo* genes after the biolistic introduction of the constructs into leaf epidermal cells¹⁵ (Supplementary Information 2). Silencing of the *HvGα* gene did not alter susceptibility to the fungus in the *Mlo* genetic background, whereas dsRNAi of *Mlo* resulted in greatly enhanced resistance (Supplementary Information 3). Similarly, application of the known membrane-permeable G-protein activators Mas-7 and cholera toxin to detached leaves did not change infection phenotypes of *Mlo* wild-type and *mlo* mutant plants (data not shown). Thus, genetic and pharmacological data provided no evidence for the involvement of heterotrimeric G proteins in the action of MLO or of the components of MLO-regulated resistance against powdery mildew¹⁶.

Similarly to a previous study⁸, bombardment of *mlo* resistant leaves with wild-type *Mlo* constructs resulted in susceptibility to powdery mildew in 73% of transformed cells owing to cell-autonomous complementation of *mlo* mutations (defined as 100%

relative susceptibility; Fig. 1a). However, bombardment of wild-type *Mlo* leaves resulted in a striking super-susceptibility phenotype in which almost all *GUS*-expressing (transformed) cells were rendered susceptible to the fungus (Fig. 1b). Bombardment with constructs containing *mlo* resistance alleles that code for stable and properly targeted proteins did not influence susceptibility, demonstrating that the super-susceptibility phenotype required the over-expression of wild-type MLO (data not shown). The increase in susceptibility after the introduction of additional MLO into wild-type *Mlo* cells indicates that suppression of the resistance response against powdery mildew by an endogenous homozygous wild-type *Mlo* gene is incomplete and that *mlo* resistance operates, albeit at lower levels, in *Mlo* genotypes.

An interaction between calmodulin (CaM) and MLO proteins was first found by screening a rice cDNA expression library in *Escherichia coli* with the use of soybean calmodulin (SCaM1 (ref. 17) conjugated to horseradish peroxidase (CaM–HRP). This identified a rice MLO homologue (GenBank accession no. AF388195), which through deletion analysis was shown to contain a calmodulin-binding domain (CaMBD) in a region of the carboxy-terminal cytoplasmic tail that is relatively conserved in sequence among MLO family members (Supplementary Information 4, and data not shown). To determine whether barley MLO also binds CaM, we fused the C-terminal MLO residues 414–435 to the C-terminus of glutathione S-transferase (GST), which itself does not bind CaM. In addition, we constructed four mutants in which single amino acids in the presumptive binding site were replaced (L420R, W423R, N422K and E429K; Supplementary Information 4). GST fusion proteins were expressed in *E. coli* (Fig. 2a, b) and subjected to CaM–HRP overlay assays¹⁷ in the presence of CaCl₂. This revealed strong binding to the wild-type sequence, greatly diminished binding or a lack of binding to the L420R and W423R mutants, unaltered binding to the E429K variant, and slightly enhanced binding to the N422K derivative (Fig. 2c). Each of the interactions was

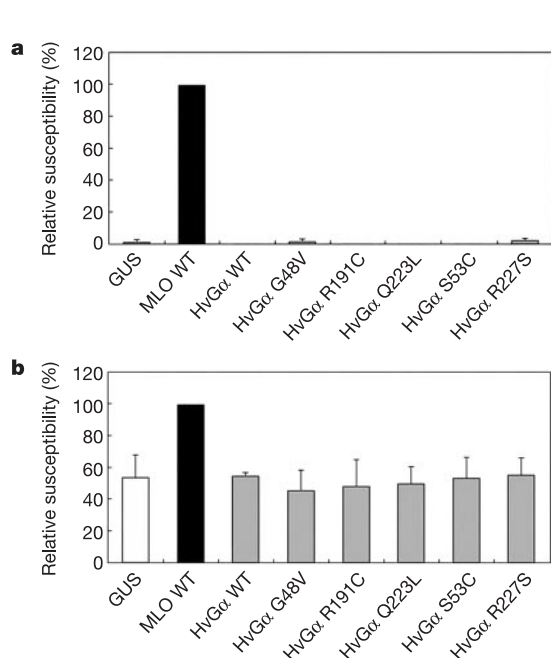


Figure 1 MLO functions independently of heterotrimeric G proteins. Resistant near-isogenic barley lines Ingrid *mlo-5* (a) and susceptible Ingrid *Mlo* (b) were either bombarded with a *GUS* reporter plasmid or bombarded together with a construct containing *Mlo*, Gα wild-type (WT) or Gα mutant variants G48V, R191C, Q223L, S53C or R227S. Leaves were then inoculated with *Bgh* and stained for *GUS* activity and fungal structures as described in Methods.

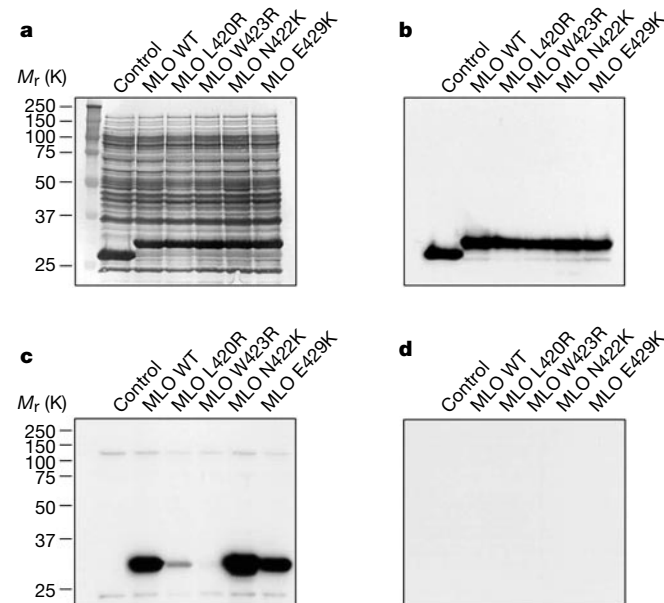


Figure 2 CaM binding to the MLO CaMBD *in vitro*. GST–CaMBD fusion proteins of either the wild-type MLO CaMBD or MLO CaMBD mutant variants L420R, W423R, N422K or E429K were expressed in *E. coli* and corresponding bacterial lysates were tested by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (a), Western blot analysis (probed with a GST-specific antiserum) (b) and a SCaM1 overlay assay (as described in Methods; CaM–HRP overlay) in the presence of 1 mM CaCl₂ (c) or 5 mM EGTA (d) (control, GST only). Qualitatively identical results were obtained with the use of SCaM4 overlay assays. The relative molecular masses (*M_r*) of marker proteins are indicated at the left.

disrupted by EGTA, indicating a requirement of Ca^{2+} for complex formation (Fig. 2d).

To test direct interactions between full-length MLO proteins and CaM in living cells, we adopted the yeast split-ubiquitin technique¹⁸. Bait constructs driving expression of barley *Mlo* and *Arabidopsis AtMlo1* and *AtMlo2* were tested for interaction with barley HvCaM3 (99% identical to each of seven *Arabidopsis* CaMs). Although we could not detect an interaction between HvMLO and HvCaM3 under the conditions tested, AtMLO1 and AtMLO2 each interacted with HvCaM3 as indicated by the growth of the corresponding yeast strains on medium containing 5-fluoro-orotic acid (5-FOA; Supplementary Information 5). To test the dependence of the interaction on the CaMBD sequence identified, we compared wild-type and AtMLO1 CaMBD substitution mutants that each showed indistinguishable bait protein stability. Whereas the single amino-acid substitutions W456R and W456Y (corresponding to W423 in barley MLO; Supplementary Information 4) did not affect growth on 5-FOA medium, the double mutant L453R/W456R showed a markedly increased 5-FOA sensitivity (Supplementary Information 5), indicative of a destabilized interaction between bait (AtMLO1) and prey (HvCaM3).

To test the significance of the barley CaM–MLO interaction *in vivo* in pathogen defence, we introduced the four CaMBD mutations examined *in vitro* into full-length *Mlo* and tested the ability of these variants to complement the *mlo-5*-null mutant (Fig. 3a). Variants L420R and W423R showed $66 \pm 5\%$ and $56 \pm 6\%$ (means $\pm$ s.d.) of wild-type *Mlo* activity, respectively; mutant E429K conferred almost full wild-type *Mlo* activity ($99 \pm 7\%$); and N422K resulted in $107 \pm 3\%$ of wild-type *Mlo* activity (that is, a mild super-susceptible phenotype). The mutants showed similar relative activities in a wild-type *Mlo* background (Fig. 3b). A second

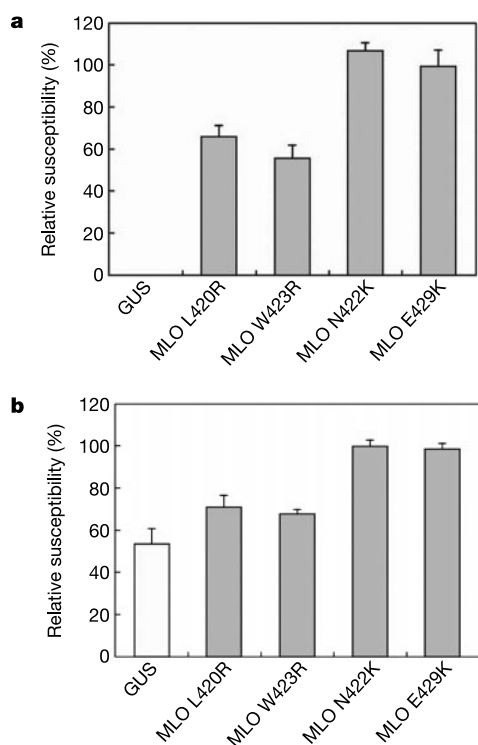


Figure 3 *In vivo* activity of CaMBD mutants. CaMBD MLO mutant activities were assessed by biolistic bombardment. Resistant (*mlo*) (a) or susceptible (*Mlo*) (b) barley genotypes were either bombarded with a GUS reporter plasmid or bombarded together with a construct containing wild-type *Mlo* or one of the *Mlo* CaMBD mutants L420R, W423R, N422K or E429K. After bombardment, leaves were inoculated with *Bgh* and stained for GUS activity and fungal structures as described in Methods.

generation of MLO CaMBD variants, including a conservative W423Y substitution (affecting the sole invariant residue of compiled MLO CaMBDs; Supplementary Information 4) as well as L420R/W423R and K431E/K432E double mutants showed, like the first set of mutants, shifts of infection phenotypes *in vivo* that reflected the shifts seen in the CaM-binding assays *in vitro* (Supplementary Information 6). To exclude the possibility that CaMBD mutations affected protein stability, we compared the accumulation of MLO and mutants L420R, W423R, W423Y and L420R/W423R *in planta* and found unaltered stability (Supplementary Information 7). The combined data from studies *in vitro* and *in vivo* therefore provide strong evidence that MLO activity is enhanced by CaM binding.

Gene silencing of *HvCaM3* by dsRNAi in an *Mlo* background halved the susceptibility seen in *Mlo* wild-type leaves (Supplementary Information 3), which is consistent with an enhancing function for CaM in MLO defence suppression. We also investigated the effects of overexpressing barley or soybean¹⁷ CaM isoforms in single epidermal cells by placing the respective genes behind the strong constitutive ubiquitin promoter. This increased susceptibility in the wild-type *Mlo* background (Fig. 4a), although to a smaller degree than bombardment with the wild-type *Mlo* construct (20% versus 44% increase in susceptibility, respectively). The super-susceptibility was dependent on wild-type CaM activity because replacement of M125, a residue known to be crucial for CaMBD binding¹⁹, with arginine did not alter susceptibility (Fig. 4b). Resistance suppression by plant CaMs required the presence of wild-type MLO because its expression in the mutant *mlo* background did not influence the resistant infection phenotype (data not shown). This supports an activator role for CaM in *Mlo*-mediated defence suppression and places CaM activity upstream of, or coincident with, the action of MLO.

A rapid increase in cytoplasmic free Ca^{2+} levels ($[\text{Ca}^{2+}]_{\text{cyt}}$) has been shown to be a common response of plant cells to pathogen challenge^{20–23}. Evidence from previous studies suggests that the pathogen-triggered Ca^{2+} fluxes have a function in triggering resistance. For example, the application of Ca^{2+} chelators or Ca^{2+} channel blockers compromises the oxidative burst that is thought to drive resistance responses at sites of attempted invasion by pathogens^{21,23}. Interestingly, our data suggest that an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ might also promote resistance suppression. CaM facilitates Ca^{2+} -dependent responses by undergoing a $[\text{Ca}^{2+}]$ -dependent coordination of four Ca^{2+} ions that alters the affinity of CaM for its target proteins¹⁹. Therefore, if a rapid increase of $[\text{Ca}^{2+}]_{\text{cyt}}$ is also a response of barley epidermal cells to barley mildew, as preliminary

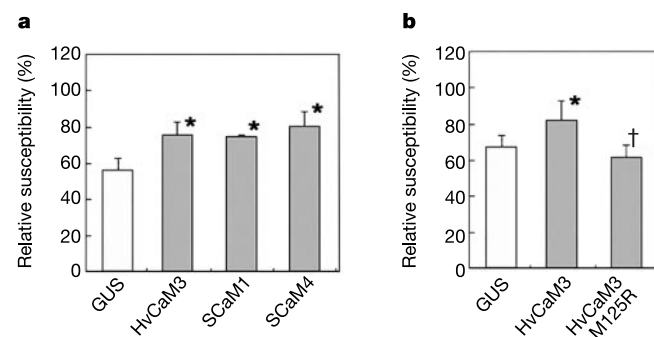


Figure 4 Single-cell expression of CaM genes. The susceptible (*Mlo*) barley genotype was co-bombarded with a GUS reporter plasmid and a construct containing barley *HvCaM3*, soybean *SCaM1* or *SCaM4* (a), or a GUS reporter plasmid and a construct containing *HvCaM3* or its mutant derivative M125R (b). After bombardment, leaves were inoculated with *Bgh* and stained for GUS activity and fungal structures as described in Methods. Asterisks indicate $P < 0.05$ compared with GUS; the dagger indicates $P < 0.05$ compared with *HvCaM3*.

Fluo-4-based Ca^{2+} indicator dye experiments indicate (data not shown), the resistance-suppressing activity of MLO might be boosted soon after pathogen challenge, through Ca^{2+} -dependent CaM binding.

Interactions of heterotrimeric G proteins with seven-transmembrane (7-TM) proteins serve as a paradigm in signal transduction processes in fungi and animals⁶. However, recent evidence suggests that some 7-TM receptors can transmit the same extracellular signal through both heterotrimeric G-protein-dependent and G-protein-independent mechanisms²⁴. Examples of the latter are the β_2 -adrenoreceptor-dependent activation of the Src kinase pathway and the metabotropic-glutamate-receptor-mediated release of Ca^{2+} in neuronal cells, each requiring the C-terminal tails of the 7-TM proteins (reviewed in ref. 24). 7-TM proteins of the opioid and the D2-dopamine receptor subclass also contain a functional CaMBD that resides in the third intracellular loop along with the G-protein-binding domain, resulting in direct binding competition²⁵. Here we show that the signalling activity of the 7-TM protein MLO is enhanced by Ca^{2+} -dependent CaM binding to a CaMBD present in the C-terminal cytoplasmic tail, and that MLO-dependent defence suppression does not require heterotrimeric G proteins.

A comparison of MLO family members revealed hypervariability in the length and sequence of the first extracellular loop⁵. This, together with the findings that mutations in the cytoplasmic CaMBD maximally compromise half of the MLO wild-type activity, is compatible with a model in which MLO integrates both intracellular and extracellular signals of host or pathogen origin to modulate defence. Conceivably, the powdery mildew fungus targets MLO to mediate defence suppression. A potential indicator of the biological role of MLO-mediated defence modulation is provided by enhanced susceptibility phenotypes of *mlo* mutants to the necrotrophic pathogen *Bipolaris sorokiniana* and the hemibiotrophic *Magnaporthe grisea*^{26,27}. Thus, the MLO modulator might ensure a balance between mutually inhibitory defence responses to different pathogens. □

Methods

Transient expression in barley epidermal cells

Effects of gene constructs on susceptibility to powdery mildew were tested by using a biolistic method for expressing constructs in single barley-leaf epidermal cells⁸. Generally, genes were driven by the maize ubiquitin promoter and were followed by the *Agrobacterium* NOS transcriptional terminator. Leaf sections were heavily inoculated with conidiospores of *Bgh* isolate K1, normally 4 h after bombardment, except in the dsRNAi experiment, in which inoculation was performed 96 h after bombardment to allow sufficient turnover of the targeted endogenous protein. At 48 h after inoculation, leaves were stained for GUS activity and for fungal structures as described previously¹⁵.

Microscopic analysis

Microscopic examination of the GUS-expressing (transformed) cells was used to determine the frequency with which challenge by powdery mildew resulted in cell-wall penetration and fungal growth as evidenced by the occurrence of fungal haustoria within the cells and epiphytic hyphae. Susceptibility obtained by bombardment with a wild-type *Mlo* gene construct was set as 100% relative susceptibility in all experiments, except the dsRNAi experiment (Supplementary Information 3), in which susceptibility obtained in cells transformed with the GUS control vector was set as 100%. Data for each construct are means and standard deviations obtained from at least three independent experiments, scoring about 100 transformed cells per experiment.

CaMBD binding assay *in vitro*

Oligonucleotide fragments coding for either the wild-type MLO CaMBD or the mutant variants shown in Fig. 2a were fused in frame to GST in the bacterial expression vector pAK-SS (a pGEX-2T derivative) and transformed into *E. coli* BL21(DE3)pLysS. Bacterial lysates were obtained by sonication, subjected to polyacrylamide gel electrophoresis (20 μg each) and stained with Coomassie brilliant blue. After Western blot transfer, the membrane was probed with a polyclonal GST-specific antiserum. A duplicate blot was probed with a soybean CaM–HRP conjugate (as described in ref. 17) in the presence of 1 mM CaCl_2 or 5 mM EGTA, and bound CaM was detected by chemiluminescence.

Yeast split ubiquitin assay

Coding regions (complementary DNAs) of barley *Mlo*, *AtMlo1* and *AtMlo2* were cloned into the vector pMet-Ste14-Cub-R-Ura3, replacing yeast *Ste14* (ref. 18). HvCaM3 cDNA was cloned into modified versions of vectors pCup-NuL-Sec62 and pCup-NuA-Sec62, replacing yeast *Sec62*. *Saccharomyces cerevisiae* strain JD53 was used for all experiments.

Interaction for each pair of proteins was tested on selective medium containing 5-FOA and on selective medium lacking uracil.

Received 3 January; accepted 5 February 2002.

- Dangl, J. L. & Jones, J. D. G. Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833 (2001).
- Shirasu, K. & Schulze-Lefert, P. Regulators of cell death in disease resistance. *Plant Mol. Biol.* **44**, 371–385 (2000).
- Jørgensen, J. H. Discovery, characterization and exploitation of *Mlo* powdery mildew resistance in barley. *Euphytica* **63**, 141–152 (1992).
- Büschges, R. *et al.* The barley *Mlo* gene: A novel control element of plant pathogen resistance. *Cell* **88**, 695–705 (1997).
- Devoto, A. *et al.* Topology, subcellular localization, and sequence diversity of the *Mlo* family in plants. *J. Biol. Chem.* **274**, 34993–35004 (1999).
- Bockaert, J. & Pin, J. P. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18**, 1723–1729 (1999).
- Ma, H. The different faces of GPA1. *Curr. Biol.* **11**, 869–871 (2001).
- Shirasu, K., Nielsen, K., Piffanelli, P., Oliver, R. & Schulze-Lefert, P. Cell-autonomous complementation of *mlo* resistance using a biolistic transient expression system. *Plant J.* **17**, 293–299 (1999).
- Winitz, S., Russell, M. & Johnson, G. L. Construction of mutant and chimeric G-protein α subunits. *Methods Enzymol.* **237**, 321–326 (1994).
- Regenfelder, E. *et al.* G proteins in *Ustilago maydis*: Transmission of multiple signals? *EMBO J.* **16**, 1934–1942 (1997).
- Aharon, G. S., Gelli, A., Snedden, W. A. & Blumwald, E. Activation of a plant plasma membrane Ca^{2+} channel by TG α 1, a heterotrimeric G protein α -subunit homologue. *FEBS Lett.* **424**, 17–21 (1998).
- Lein, W. & Saalbach, G. Cloning and direct G-protein regulation of phospholipase D from tobacco. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **1530**, 172–183 (2001).
- Slepak, V. Z. *et al.* Random mutagenesis of G protein α subunit $G_{i\alpha}$. Mutations altering nucleotide binding. *J. Biol. Chem.* **268**, 21889–21894 (1993).
- Apanovitch, D. M., Iiri, T., Karasawa, T., Bourne, H. R. & Dohlman, H. G. Second site suppressor mutations of a GTPase-deficient G-protein α -subunit. Selective inhibition of G $\beta\gamma$ -mediated signaling. *J. Biol. Chem.* **273**, 28597–28602 (1998).
- Schweizer, P., Pokorný, J., Schulze-Lefert, P. & Dudler, R. Double-stranded RNA interferes with gene function at the single-cell level in cereals. *Plant J.* **24**, 895–903 (2000).
- Peterhansel, C., Freialdenhoven, A., Kurth, J., Kolsch, R. & Schulze-Lefert, P. Interaction analyses of genes required for resistance responses to powdery mildew in barley reveal distinct pathways leading to leaf cell death. *Plant Cell* **9**, 1397–1409 (1997).
- Lee, S. H. *et al.* Differential activation of NAD kinase by plant calmodulin isoforms. The critical role of domain I. *J. Biol. Chem.* **272**, 9252–9259 (1997).
- Stagljar, I., Korostensky, C., Johnsson, N. & Heesen, S. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 5187–5192 (1998).
- Chin, D. & Means, A. R. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* **10**, 322–328 (2000).
- Xu, H. & Heath, M. C. Role of calcium in signal transduction during the hypersensitive response caused by basidiospore-derived infection of the cowpea rust fungus. *Plant Cell* **10**, 585–597 (1998).
- Grant, M. *et al.* The *RPM1* plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. *Plant J.* **23**, 441–450 (2000).
- Jabs, T., Tschöpe, M., Colling, C., Hahlbrock, K. & Scheel, D. Elicitor-stimulated ion fluxes and O_2 from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley. *Proc. Natl Acad. Sci. USA* **94**, 4800–4805 (1997).
- Blume, B., Nürnberg, T., Nass, N. & Schell, D. Receptor-mediated increase in cytoplasmic free calcium required for activation of pathogen defense in parsley. *Plant Cell* **12**, 1425–1440 (2000).
- Brzostowski, J. A. & Kimmel, A. R. Signaling at zero G: G-protein-independent functions for 7-TM receptors. *Trends Biochem. Sci.* **26**, 291–297 (2001).
- Bofill-Cardona, E. *et al.* Binding of calmodulin to the D₂-dopamine receptor reduces receptor signaling by arresting the G protein activation switch. *J. Biol. Chem.* **275**, 32672–32680 (2000).
- Jarosch, B., Kogel, K. H. & Schaffrath, U. The ambivalence of the barley *Mlo* locus: Mutations conferring resistance against powdery mildew (*Blumeria graminis* f.sp. *hordei*) enhance susceptibility to the rice blast fungus *Magnaporthe grisea*. *Mol. Plant–Microbe Interact.* **12**, 508–514 (1999).
- Kumar, J., Hükelhoven, R., Beckhove, U., Nagarajan, S. & Kogel, K. H. A compromised *Mlo* pathway affects the response of barley to the necrotrophic fungus *Bipolaris sorokiniana* (Teleomorph: *Cochliobolus sativus*) and its toxins. *Phytopathology* **91**, 127–133 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank N. Collins for critical comments on the manuscript, G. Jach for providing RFP vector pGJ1425, N. Johnsson for making available split-ubiquitin plasmids, B. Schauf for technical assistance, and J. Glazebrook at Syngenta's TMRI for her help in inspecting the rice genome for genes coding for heterotrimeric G-protein α subunits. M.C.K. and R.P. contributed equally to the work and are listed in alphabetical order in the author list. This work was supported by grants from the Max-Planck Society and the Gatsby Charitable Organization to P.S.-L., and grants from Korean Ministry of Science and Technology NRL, G7 and Crop Functional Genomic Centre of New Frontier to M.J.C.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.S.-L. (e-mail: schlef@mpiz-koeln.mpg.de) or M.J.C. (e-mail: mjcho@nongae.gsnu.ac.kr).

A single motif responsible for ubiquitin recognition and monoubiquitination in endocytic proteins

Simona Polo*, Sara Sigismund*, Mario Faretta*, Monica Guidi*, Maria Rosaria Capua*, Giovanna Bossi*, Hong Chen†, Pietro De Camilli‡ & Pier Paolo Di Fiore*‡

* Department of Experimental Oncology, European Institute of Oncology, Via Ripamonti 435, 20141, Milan, Italy

† Department of Cell Biology and Howard Hughes Medical Institute, Yale University School of Medicine, 295 Congress Avenue, New Haven, Connecticut 06510, USA

‡ IFOM, The FIRC Institute for Molecular Oncology, 20134 Milan; and University of Milan, Medical School, 20122 Milan, Italy

Ubiquitination is a post-translation modification in which ubiquitin chains or single ubiquitin molecules are appended to target proteins, giving rise to poly- or monoubiquitination, respectively^{1–4}. Polyubiquitination targets proteins for destruction by the proteasome. The role of monoubiquitination is less understood, although a function in membrane trafficking is emerging, at least in yeast^{1,3,5}. Here we report that a short amino-acid stretch at the carboxy-termini of the monoubiquitinated endocytic proteins Eps15 and eps15R is indispensable for their monoubiquitination. A similar sequence, also required for this modification, is found in other cytosolic endocytic proteins, such as epsins and Hrs. These sequences comprise a protein motif, UIM (ref. 6), which has been proposed to bind to ubiquitin. We confirm this for the UIMs of eps15, eps15R, epsins and Hrs. Thus, the same motif in several endocytic proteins is responsible for ubiquitin recognition and monoubiquitination. Our results predict the existence of a UIM:ubiquitin-based intracellular network. Eps15/eps15R, epsins and Hrs may function as adaptors between ubiquitinated membrane cargo and either the clathrin coat or other endocytic scaffolds. In addition, through their own ubiquitination, they may further contribute to the amplification of this network in the endocytic pathway.

Eps15 is ubiquitinated after the activation of the epidermal growth factor (EGF) receptor (Fig. 1a, and ref. 7). This event is associated with an increase in the apparent relative molecular mass, M_r , of ~8K, compatible with the molecular mass of a single ubiquitin molecule (Fig. 1a, right, and ref. 7). The related protein eps15R, for which several isoforms have been described⁸, is also ubiquitinated following EGF receptor activation (Fig. 1a). The M_r of ubiquitinated eps15R is compatible with monoubiquitination of the isoforms migrating in the M_r ~125 and ~108K region, whereas the 76K form is not modified (Fig. 1a).

Eps15 and eps15R share the same modular architecture⁹ comprising (from amino to carboxy terminus) three EH domains, a coiled-coil region and a C-terminal region, which binds to the clathrin adaptor complex AP2 (Fig. 1b). We mapped the regions that are important for ubiquitination by employing an ubiquitination assay and various fragments of the proteins as substrates. The C-terminal region, glutathione S-transferase (GST)–Cter, of both proteins contained all the determinants necessary for ubiquitination (Fig. 1b). Eps15 and eps15R share little similarity in their C-terminal regions, with the notable exception of a stretch of about 40 amino acids at their ends¹⁰. We engineered deletions of the last 23 amino acids of eps15, either in the context of the holoprotein or of its C-terminal region, and tested their ubiquitination *in vitro*. We could not detect ubiquitination of the truncated proteins (Fig. 1c).

Next, a series of mutants in the C-terminal region of eps15 were

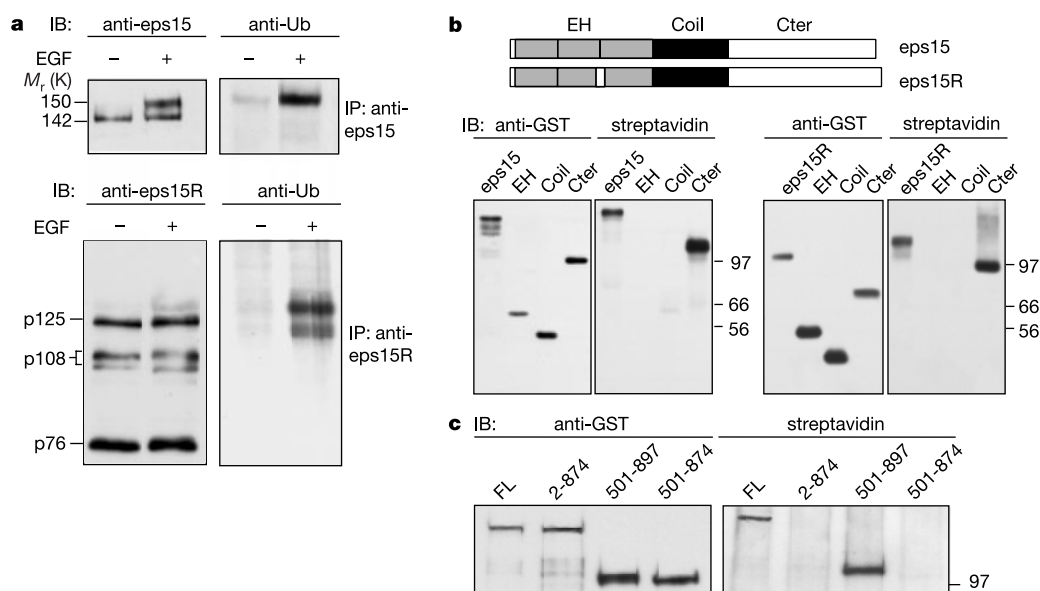


Figure 1 A motif in eps15 and eps15R responsible for monoubiquitination. **a**, B82L cells were either treated with EGF (+) or mock-treated (–). Total cellular lysates (4 mg) were immunoprecipitated (IP) and then immunoblotted (IB) with the indicated antibodies. The positions of eps15, monoubiquitinated eps15, and of the various isoforms of eps15R are indicated on the left. **b**, Top, scheme of the modular structure of eps15 and eps15R: EH domains, coiled-coil region and C-terminal region. Bottom, GST fusion proteins (0.3 μg) encompassing full length or fragments of eps15 (left panels) or eps15R (right panels) were

subjected to an *in vitro* ubiquitination assay, using biotinylated ubiquitin (Methods). Detection was performed with horseradish-peroxidase-conjugated streptavidin or anti-GST, as indicated. **c**, GST fusion proteins, either full-length eps15 (FL) or fragments thereof (as indicated by the amino-acid positions above each lane) were subjected to the *in vitro* ubiquitination (Ub) assay, as in **b**. In **b** and **c**, relative molecular mass (M_r) markers are shown on the right.

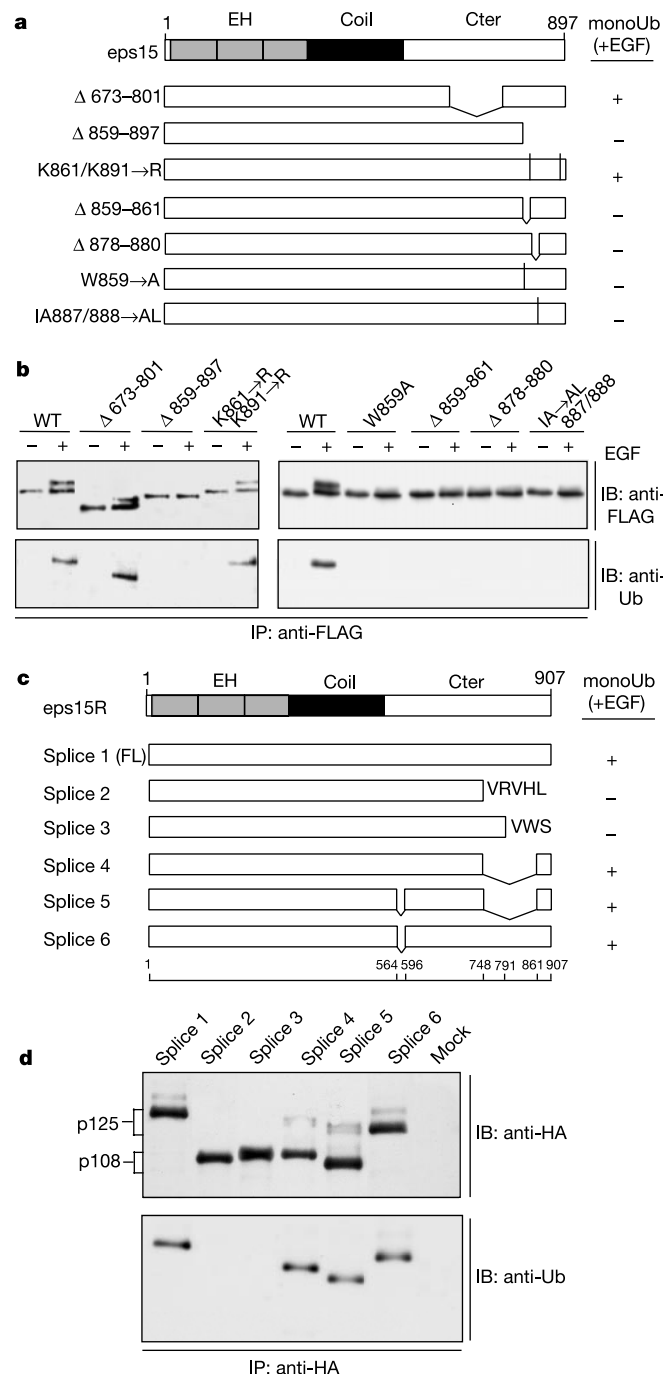


Figure 2 *In vivo* mapping of the motif required for monoubiquitination of eps15 and eps15R. **a**, Scheme of the eps15 mutants. The ubiquitination of the mutants (in the presence of EGF) is also reported (– or +). **b**, Mutants shown in **a** were engineered into eukaryotic expression FLAG-vectors, and transfected into B82L cells (WT, wild-type eps15). Total cellular lysates (5 mg) from cells treated with EGF (+) or mock-treated (–), were immunoprecipitated with anti-FLAG antibodies, followed by immunoblot with the indicated antibodies. **c**, Six differently spliced isoforms of eps15R were isolated (see Supplementary Information) and their conceptually translated products are shown. For isoforms 2 and 3, the sequence contains five or three extra C-terminal amino acids, respectively, resulting from the alternative splicing events. The ubiquitination of the various isoforms (in the presence of EGF) is also reported (– or +). **d**, HA-tagged versions of the six isoforms, shown in **c**, were individually expressed in COS cells and the cells were treated with EGF. Total cellular lysates (5 mg) were immunoprecipitated with the anti-HA antibody, followed by immunoblot with the indicated antibodies. The apparent M_r of the various isoforms (top) was compatible (taking into account the HA epitope) with that of the endogenously expressed isoforms that migrate in the M_r ~125K and ~108K regions (compare with Fig. 1a).

engineered in the context of the holoprotein (Fig. 2a), and their ubiquitination was assayed *in vivo*. As expected, a truncation of the last 39 amino acids (Δ 859–897) yielded a protein that could no longer be ubiquitinated following EGF treatment (Fig. 2b). Small interstitial deletions or mutations in the same region (Δ 859–861, Δ 878–880, W859 → A, and IA887/888 → AL) also yielded ubiquitination-impaired mutants (Fig. 2b). Conversely, a longer deletion in the area immediately upstream (Δ 673–801), which removed the AP2-binding region^{11,12}, behaved indistinguishably from the wild type, confirming the boundaries of the region critical for ubiquitination (Fig. 2b). Two lysines, which could represent acceptor sites for ubiquitin, are present in the last 39 amino acids of eps15. We mutagenized these to arginine individually (not shown) or simultaneously (Fig. 2a). These mutants were ubiquitinated *in vivo* as efficiently as wild-type eps15 (not shown and Fig. 2b), indicating that we did not map the lysine(s) that function as acceptor(s) of ubiquitin, but rather a region important for other aspects of the process; probably recognition by ubiquitin ligase(s).

Eps15R provided us with the opportunity of testing the relevance for ubiquitination of the C-terminal region in a physiological setting. We isolated six splicing variants of eps15R (Supplementary Information), which correspond to various isoforms migrating in the M_r 125 and 108K regions (Fig. 2c, d). When individually expressed in COS (African green monkey) cells, only those isoforms retaining the last coding exon, corresponding to the C-terminal 46

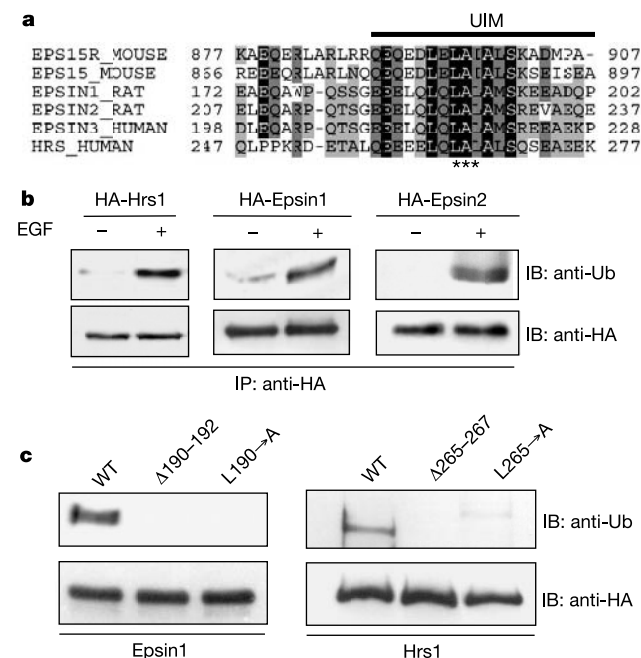


Figure 3 Monoubiquitination of other proteins displaying the 'monoubiquitination' motif. **a**, Alignment of regions of local homology among epsins, Hrs and the motif responsible for monoubiquitination of eps15 and eps15R. The alignment was drawn using the Belvu program (<http://www.cgr.ki.se/cgr/groups/sonnhammer/Belvu.html>) with the default colouring method. Increasing intensity of shading corresponds to increasing conservation. Amino-acid positions are also indicated. The boundaries of UIM, as identified in ref. 6, are indicated by a bar. Asterisks identify the amino acids mutagenized in the experiments shown in **c**. **b**, HA-tagged versions of epsin1, epsin2 and Hrs were transfected in COS cells, which were subsequently either treated with EGF (+) or mock-treated (–). **b**, **c**, Total cellular lysates (5 mg) were immunoprecipitated with anti-HA antibodies, followed by immunoblot with the indicated antibodies. **c**, Small deletion or point mutants (the relevant amino acids are indicated by their positions in the holoproteins, and also by asterisks in **a**), in the UIM of epsin1 and Hrs, in HA-tagged versions, were transfected in COS cells, which were subsequently treated with EGF.

amino acids, displayed ubiquitination (Fig. 2c, d). We note that the M_r 76K isoform, whose biogenesis is unknown but lacks the entire C-terminal region downstream of the coiled-coil (M. Vecchi, S. Sigismund and P. P. Di Fiore, unpublished work), was not ubiquitinated (Fig. 1a).

We searched databases for regions of local homology to the region critical for ubiquitination of eps15 and eps15R. Several proteins displayed significant scores; the closest homologies were detected in proteins that, like eps15 and eps15R, are proposed to function in membrane traffic, such as epsin¹³ and Hrs¹⁴ (Fig. 3a). We predicted that these proteins would be substrates for monoubiquitination *in vivo* as well. Indeed, haemagglutinin (HA)-tagged versions of all these proteins were monoubiquitinated in EGF-treated cells (Fig. 3b). Furthermore, epsin1 and Hrs harbouring mutations in this region were no longer ubiquitinated after EGF treatment (Fig. 3c).

The identified motif, shared by these proteins, overlaps with the recently described UIM present in two copies in the S5a proteasome subunit, as well as in a variable number of copies in several other proteins⁶. Because the two UIMs of S5a are responsible for binding to ubiquitin¹⁵, it was proposed that all UIM-containing proteins have ubiquitin-binding properties, but this possibility was not tested experimentally. Indeed, we found that the GST–Cter of eps15 and eps15R bound to numerous ubiquitinated proteins in total cellular lysates (Fig. 4b).

The last 40 amino acids of eps15 contain two UIMs (defined according to ref. 6, Fig. 4a). Mutations in the first UIM (Δ 859–861 and W859 $\rightarrow$ A) abrogated the ability of eps15 to be monoubiquitinated but did not alter the ability of the protein to bind to ubiquitin-containing proteins, whereas mutations in the second UIM (Δ 878–880) did both (Fig. 4b). The double lysine mutant

displayed unaltered ubiquitin-binding properties (Fig. 4b). Isolated UIMs from eps15, epsin1 and Hrs displayed binding to ubiquitin-containing proteins (Fig. 4c). They could also bind to monoubiquitin (Fig. 4d), albeit with relatively low affinity (Fig. 4d, see also Supplementary Information). The GST–S5a protein was also able to bind to monoubiquitin with an efficiency comparable to that of the analysed endocytic proteins (Supplementary Information). We note that all of the tested UIMs displayed a preference for polyubiquitin chains, with respect to monoubiquitin, as expected from previous findings^{15–18} (Supplementary Information). We conclude that the UIM is endowed with the dual property of binding to ubiquitin and being indispensable for monoubiquitination, at least in the analysed proteins.

Ubiquitination is a multistep process involving a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), which accepts ubiquitin from E1, and finally a ubiquitin ligase (E3), which catalyses the transfer of ubiquitin from E2 to the substrate². E3s of the HECT subfamily¹⁹, such as Nedd4 (ref. 20), form a thiol-ester intermediate with ubiquitin, as part of this process. Work in yeast suggested a key role for Rsp5 (the yeast homologue of the mammalian Nedd4 family of HECTs) in the monoubiquitination of endocytic proteins^{21–23}. In addition, genetic interaction was detected between the genes coding for Rsp5p and Ede1p, the yeast homologue of eps15 (ref. 24). Thus, HECT proteins, and in particular Nedd4, might be involved in monoubiquitination of endocytic proteins. Indeed, overexpression of Nedd4 *in vivo* led to increased monoubiquitination of eps15 (Fig. 5a). Accordingly, in an *in vitro* reconstituted system containing solely E1, E2 (UbcH5B, ref. 25) and Nedd4 (as the E3), Nedd4 was able to monoubiquitinate eps15 (Fig. 5b). Because Nedd4 does not

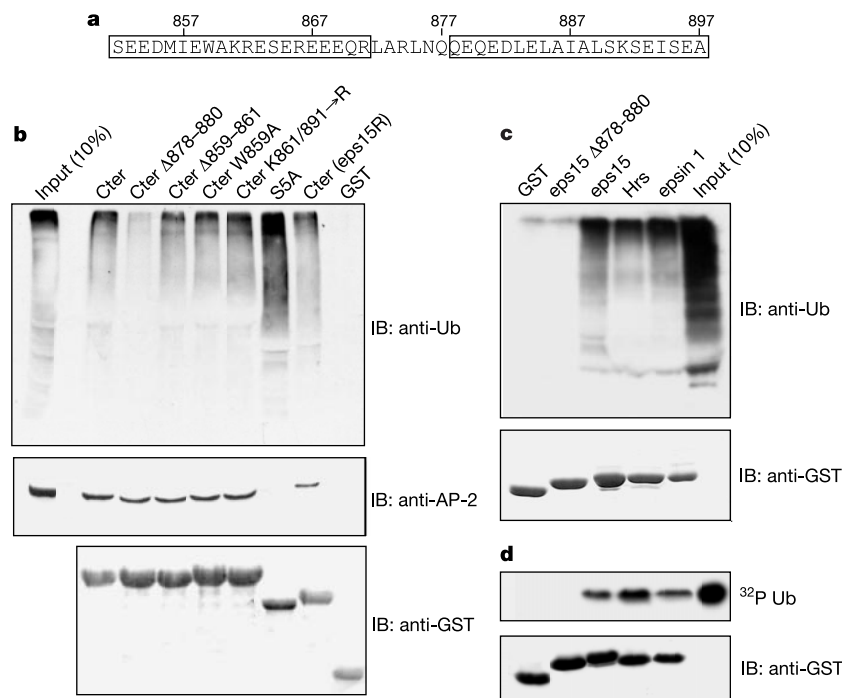


Figure 4 Ubiquitin-binding properties of UIMs. **a**, The sequence of the last 46 amino acids of eps15 is shown, with the two UIMs boxed in. Amino-acid positions are also shown. **b**, Top, the indicated GST fusion proteins (15 μ g) were used to pull down ubiquitin-containing proteins from total cell lysates of COS cells (500 μ g), followed by immunoblot with anti-ubiquitin antibodies. The GST fusions were: the C-terminal region of eps15 (Cter, amino acids 501–897), the C-terminal region of eps15R (Cter–eps15R, amino acids 597–907), the full-length S5a (S5a) and the indicated eps15 mutants all engineered in the context of GST–Cter. The same blot was redecorated with anti- α -adaptin (middle) as a positive control, and with anti-GST (bottom) to show comparable

loading. **c**, Isolated UIMs from eps15 (amino acids 862–897), epsin1 (amino acids 182–227) and Hrs (amino acids 257–277), engineered as GST fusion proteins (15 μ g), were used to pull down ubiquitin-containing proteins from total cell lysates of COS cells (500 μ g), followed by immunoblot with anti-ubiquitin antibodies. The eps15 Δ 878–880 construct is a GST fusion protein harbouring amino acids 862–897 of eps15, with an internal deletion of amino acids 878–880, and was used as a negative control. The same blot was redecorated with anti-GST (bottom) to show comparable loading. **d**, Radioactive monomeric ubiquitin (Methods) was used in pulldown experiments with the same UIMs as in **c**. Lanes are as in **c**.

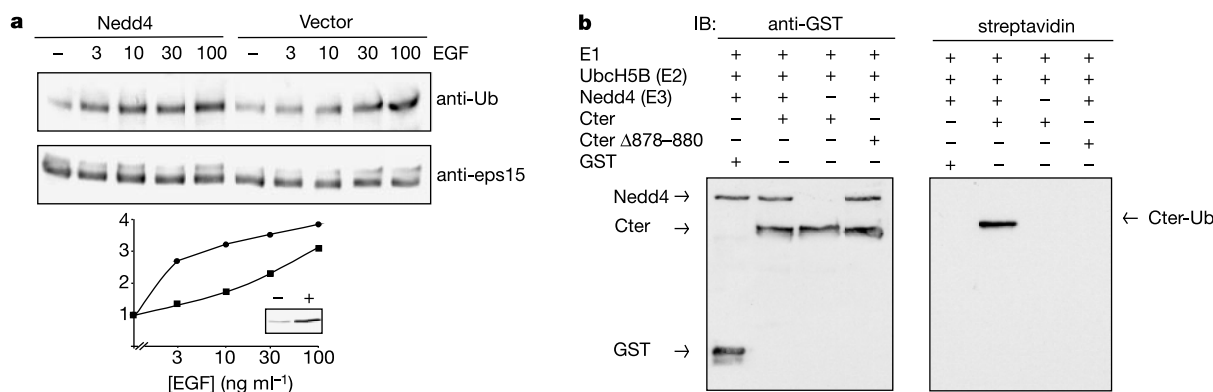


Figure 5 The UIM of eps15 is necessary for monoubiquitination by Nedd4. **a**, Nedd4 monoubiquitinates eps15 *in vivo*. Left, an expression vector for Nedd4, or a control vector, were transfected in COS cells. Cells were treated with EGF at the indicated concentrations for 10 min. Lysates were immunoprecipitated with anti-eps15, followed by immunoblotting with anti-ubiquitin (anti-Ub) (top) or anti-eps15 (bottom). Right, a densitometric scanning was performed using the NIH IMAGE software. Specific ubiquitination signals were normalized for the actual loaded protein. Signals are reported as fold increase relative to the signal in the (-) lanes. Circles, Nedd4 transfectants; squares, control vector transfectants (in the inset the levels of overexpression of Nedd4,

- and + transfectants, are shown). The calculated half-maximal inhibitory concentration, IC_{50} , for EGF was 1.7 ng ml^{-1} in the presence of overexpressed Nedd4 and 10.0 ng ml^{-1} in the control. **b**, Nedd4 monoubiquitinates eps15, but not a UIM-deleted mutant, *in vitro*. GST fusion proteins ($0.3 \mu\text{g}$) encompassing the C-terminal fragment (amino acids 501–897) of eps15, or a C-terminal fragment bearing a deletion in the second UIM ($\Delta 878-880$), were subjected to an *in vitro* ubiquitination assay using biotinylated ubiquitin and the purified enzymes indicated (Methods). Detection was with horseradish-peroxidase-conjugated streptavidin or anti-GST, as indicated.

bind stably to eps15 (Supplementary Information), this provided us with the opportunity to test whether the two properties of the UIM, that is, interaction with ubiquitin and indispensability for monoubiquitination, are linked. As shown in Fig. 5b, Nedd4 was able to monoubiquitinate eps15, but not a UIM-deleted mutant ($\Delta 878-880$), thus functionally connecting the two properties.

Our results demonstrate that ubiquitin binding is a general property of the UIM. This motif is present in a heterogeneous class of proteins that includes endocytic proteins, signal transducers, ubiquitin ligases and isopeptidases, protein and lipid kinases and others⁶. Many proteins are modified in the cell by ubiquitination, so we predict the existence of an extensive network of interactions, which could influence many aspects of cellular physiology. Of particular interest is the possibility that UIM-containing endocytic proteins might be recruited to the ubiquitinated tails of membrane proteins, for example plasma membrane receptors. Given their additional interactions with components of the clathrin coat and/or other membrane-associated scaffolds, these endocytic proteins may function as adaptors in the recruitment of ubiquitinated membrane cargo along the endocytic pathway. Our results also show that some UIM-containing proteins can be monoubiquitinated in a signalling-dependent manner. Thus, they could further assemble other UIM-containing proteins in receptor-bound macromolecular complexes and contribute to events leading to intracellular signal transduction or to sorting of receptors in endocytic vesicles. UIMs prefer polyubiquitin chains in *in vitro* binding assays (this work and refs 15,16,17,18), but whether this reflects a similar preference *in vivo* remains unresolved. However, we note that additional direct or indirect interactions of UIM-containing proteins with their ubiquitinated partners or with the surrounding lipid bilayer (for example, epsin and Hrs bind to phosphoinositides²⁶⁻²⁹) might increase overall affinity and determine specificity.

Another compelling feature of the UIM, in the analysed subset of proteins, is its absolute requirement for their monoubiquitination, a property functionally linked to its ubiquitin-binding ability. One possible interpretation of these results is that recognition between E3 and some UIM-containing proteins is mediated, at least in part, by the ubiquitin present in the thiol-ester intermediate of the E3. Also in this case, the specificity of recognition might be determined by additional interactions. In the case of eps15, the surface responsible for these additional, specificity-determining interactions

might coincide with the first UIM, which is indispensable for monoubiquitination, but does not bind to ubiquitin (Supplementary Information). After ubiquitin transfer to the substrate, the interaction between E3 and the substrate would be dissolved, preventing further catalysis and yielding a monoubiquitinated product. Alternatively, after catalysis, the UIM could contract an intramolecular interaction with the substrate-bound ubiquitin, preventing its further use for chain assembly. This model may provide a mechanistic explanation of why some proteins are monoubiquitinated, but awaits experimental confirmation. Future studies could also examine the interplay between the requirement for UIMs in the ubiquitination of proteins which harbour them, and their ability to bind to other ubiquitinated proteins. These two functions may be mutually exclusive, or the same UIM could mediate both functions, at different steps of the endocytic pathway. □

Methods

Expression vectors and antibodies

HA-, FLAG-, or GST-tagged full-length mouse eps15, mouse eps15R, rat epsin1, rat epsin2, human Hrs, and human S5a and mutants thereof were engineered by recombinant polymerase chain reaction (PCR), or by endonuclease digestion, followed by subcloning into the pcDNA3.1/FLAG or pGEX-4T2 (Pharmacia) vectors. All constructs were verified by DNA sequencing. Further details are available upon request.

Antibodies used were: anti-HA (12CA5, Babco), anti-FLAG (M2, Sigma), anti- α -adaptin (A4325, Sigma) and anti-ubiquitin (MMS-258, Babco) monoclonal antibodies, monoclonal anti-eps15, and an affinity-purified anti-eps15R polyclonal serum. Immunoprecipitation and immunoblotting and *in vitro* bindings were performed as described⁸. Immunoblotting with anti-ubiquitin antibodies was performed as described³⁰.

The His-tagged ubiquitin used in Fig. 4b, engineered to contain a protein kinase recognition site (LRRASV), was purified onto nickel beads and labelled with ^{32}P by incubating [γ - ^{32}P] ATP and cAMP kinase (Sigma) at 37°C for 30 min. Where indicated, treatment with EGF (17 nM) was performed at 37°C for 10 min.

Ubiquitination assays

For standard ubiquitination assays, reaction mixtures containing $0.3 \mu\text{g}$ of purified GST fusion proteins and $100 \mu\text{g}$ of lysate of B82L cells (prepared by three cycles of freeze-thawing) were incubated at 30°C for 1 h in the presence of $6 \mu\text{g}$ of biotinylated ubiquitin, in Ub buffer (25 mM Tris-HCl pH 7.6, 5 mM MgCl_2 , 100 mM NaCl, $1 \mu\text{M}$ DTT, $2 \mu\text{M}$ ATP). Samples were washed four times and resolved by SDS-PAGE, and detected by decoration with horseradish-peroxidase-conjugated streptavidin.

For ubiquitination assays with purified enzymes, reactions were performed as described above except for the replacement of B82L cell lysate with purified enzymes (100 ng of E1, 150 ng of purified His-tagged UbcH5B, and 400 ng of GST-Nedd4).

Received 28 December 2001; accepted 6 February 2002.

- Hicke, L. A new ticket for entry into budding vesicles—ubiquitin. *Cell* **106**, 527–530 (2001).
- Weissman, A. M. Themes and variations on ubiquitination. *Nature Rev. Mol. Cell Biol.* **2**, 169–178 (2001).
- Hicke, L. Protein regulation by monoubiquitin. *Nature Rev. Mol. Cell Biol.* **2**, 195–201 (2001).
- Laney, J. D. & Hochstrasser, M. Substrate targeting in the ubiquitin system. *Cell* **97**, 427–430 (1999).
- Katzmann, D. J., Babst, M. & Emr, S. D. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* **106**, 145–155 (2001).
- Hofmann, K. & Falquet, L. A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* **26**, 347–350 (2001).
- van Delft, S., Govers, R., Strous, G. J., Verkleij, A. J. & van Bergen en Henegouwen, P. M. Epidermal growth factor induces ubiquitination of Eps15. *J. Biol. Chem.* **272**, 14013–14016 (1997).
- Coda, L. *et al.* Eps15R is a tyrosine kinase substrate with characteristics of a docking protein possibly involved in coated pits-mediated internalization. *J. Biol. Chem.* **273**, 3003–3012 (1998).
- Santolini, E., Salcini, A. E., Kay, B. K., Yamabhai, M. & Di Fiore, P. P. The EH network. *Exp. Cell Res.* **253**, 186–209 (1999).
- Wong, W. T. *et al.* A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution. *Proc. Natl Acad. Sci. USA* **92**, 9530–9534 (1995).
- Benmerah, A., Begue, B., Dautry-Varsat, A. & Cerf-Bensussan, N. The ear of alpha-adaptin interacts with the COOH-terminal domain of the Eps15 protein. *J. Biol. Chem.* **271**, 12111–12116 (1996).
- Iannolo, G. *et al.* Mapping of the molecular determinants involved in the interaction between eps15 and AP-2. *Cancer Res.* **57**, 240–245 (1997).
- Chen, H. *et al.* Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. *Nature* **394**, 793–797 (1998).
- Komada, M. & Kitamura, N. Growth factor-induced tyrosine phosphorylation of Hrs, a novel 115-kilodalton protein with a structurally conserved putative zinc finger domain. *Mol. Cell Biol.* **15**, 6213–6221 (1995).
- Young, P., Deveraux, Q., Beal, R. E., Pickart, C. M. & Rechsteiner, M. Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. *J. Biol. Chem.* **273**, 5461–5467 (1998).
- van Nocker, S. *et al.* The multiubiquitin-chain-binding protein Mcb1 is a component of the 26S proteasome in *Saccharomyces cerevisiae* and plays a nonessential, substrate-specific role in protein turnover. *Mol. Cell Biol.* **16**, 6020–6028 (1996).
- Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M. & Pickart, C. Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl Acad. Sci. USA* **93**, 861–866 (1996).
- Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. A 26S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **269**, 7059–7061 (1994).
- Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc. Natl Acad. Sci. USA* **92**, 2563–2567 (1995).
- Staub, O. *et al.* WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na⁺ channel deleted in Liddle's syndrome. *EMBO J.* **15**, 2371–2380 (1996).
- Wang, G. *et al.* Localization of the Rsp5p ubiquitin-protein ligase at multiple sites within the endocytic pathway. *Mol. Cell Biol.* **21**, 3564–3575 (2001).
- Dunn, R. & Hicke, L. Multiple roles for Rsp5p-dependent ubiquitination at the internalization step of endocytosis. *J. Biol. Chem.* **276**, 25974–25981 (2001).
- Dunn, R. & Hicke, L. Domains of the Rsp5p ubiquitin-protein ligase required for receptor-mediated and fluid-phase endocytosis. *Mol. Biol. Cell* **12**, 421–435 (2001).
- Gagny, B. *et al.* A novel EH domain protein of *Saccharomyces cerevisiae*, Edelp, involved in endocytosis. *J. Cell Sci.* **113**, 3309–3319 (2000).
- Jensen, J. P., Bates, P. W., Yang, M., Vierstra, R. D. & Weissman, A. M. Identification of a family of closely related human ubiquitin conjugating enzymes. *J. Biol. Chem.* **270**, 30408–30414 (1995).
- Itoh, T. *et al.* Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* **291**, 1047–1051 (2001).
- Patki, V., Lawe, D. C., Corvera, A. S., Virbasius, J. V. & Chawla, A. A functional PtdIns(3)P-binding motif. *Nature* **394**, 433–434 (1998).
- Burd, C. G. & Emr, S. D. Phosphatidylinositol(3)-phosphate signaling mediated by specific binding to RING FYVE domains. *Mol. Cell* **2**, 157–162 (1998).
- Gaullier, J. M. *et al.* FYVE fingers bind PtdIns(3)P. *Nature* **394**, 432–433 (1998).
- Avantaggiati, M. L. *et al.* The SV40 large T antigen and adenovirus E1a oncoproteins interact with distinct isoforms of the transcriptional co-activator, p300. *EMBO J.* **15**, 2236–2248 (1996).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank H. Asao and A. Bean for Hrs constructs, I. Berdo for E1 purification, G. F. Draetta for UbcH5B, Nedd4 and His-Ub constructs, M. Sudol for γ EnaC constructs, and G. Gill for B82L cells. We thank A. Musacchio for comments that improved the manuscript. We also thank S. Confalonieri for assistance with bioinformatics. This work was supported by grants from the Associazione Italiana Ricerca sul Cancro, the Telethon Foundation, the CNR (Progetto Biotecnologie), and the EC (V framework).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.P.D.F. (e-mail: pdfiore@ieo.it).

Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine

Tobias Krojer^{*†}, Marta Garrido-Franco^{*}, Robert Huber^{*}, Michael Ehrmann[†] & Tim Clausen^{*}

^{*} Max-Planck-Institut für Biochemie, Abteilung Strukturforschung, Am Klopferspitz 18A, 82152 Martinsried, Germany

[†] Cardiff University, School of Biosciences, Museum Avenue, Cardiff CF10 3US, UK

Molecular chaperones and proteases monitor the folded state of other proteins. In addition to recognizing non-native conformations, these quality control factors distinguish substrates that can be refolded from those that need to be degraded¹. To investigate the molecular basis of this process, we have solved the crystal structure of DegP (also known as HtrA), a widely conserved heat shock protein that combines refolding and proteolytic activities². The DegP hexamer is formed by staggered association of trimeric rings. The proteolytic sites are located in a central cavity that is only accessible laterally. The mobile side-walls are constructed by twelve PDZ domains, which mediate the opening and closing of the particle and probably the initial binding of substrate. The inner cavity is lined by several hydrophobic patches that may act as docking sites for unfolded polypeptides. In the chaperone conformation, the protease domain of DegP exists in an inactive state, in which substrate binding in addition to catalysis is abolished.

Problems associated with protein folding need to be addressed by all living cells. Molecular chaperones and proteases control the folded state of proteins by recognizing hydrophobic stretches of polypeptides that become exposed by misfolding or unfolding. Molecular chaperones either simply bind substrates to prevent aggregation, or they actively assist in refolding. If attempts at refolding fail, irreversibly damaged proteins are degraded by proteases¹. Despite extensive research into protease and chaperone mechanisms, the control of this editing function is poorly understood. While most factors involved in protein quality control are ATP-dependent heat shock proteins³, DegP fulfils this role without consuming ATP⁴. DegP is a widely conserved extracytoplasmic protein that switches from chaperone to protease in a temperature-dependent manner², the protease activity being most apparent at elevated temperatures. Prokaryotic DegP proteins have been attributed to the tolerance against various folding stresses as well as to pathogenicity⁵. Human homologues are believed to be involved in arthritis, cell growth, unfolded protein response and apoptosis^{6–8}. Our described structural study of *Escherichia coli* DegP is a start towards dissecting the molecular mechanism of this quality control factor, providing the required stereochemical framework for further genetic, biochemical, and biophysical studies. The unique structural organization of the DegP hexamer indicates a new type of protease-chaperone machine.

Native DegP undergoes slow self cleavage. Thus we used a proteolytically inactive mutant for structural determination, in which the active site serine was replaced by alanine (S210A). After successful crystallization at room temperature, the hexagonal crystals were transferred to 4 °C, thereby locking the protein in the 'chaperone conformation'. In the asymmetric unit of the crystals, two DegP molecules were observed, which we refer to here as A and B. The DegP monomer can be divided into three functionally distinct domains, namely a protease (residues 1–259) and two PDZ domains (PDZ1, residues 260–358; PDZ2, residues 359–448) (Fig. 1a). The previously proposed amino-terminal domain⁵ con-

tributes to the protease fold. Part of this N-terminal segment, the 'Q-linker'⁹ (residues 55–79), was too flexible to be traced in the electron density. Similar to other members of the trypsin family, the protease domain of DegP has two perpendicular β -barrel lobes with a carboxy-terminal helix. The catalytic triad is located in the crevice between the two lobes. The overall fold of the two PDZ domains is similar to other PDZ domains of known structure^{10,11}; however, PDZ1 has β 13 and α f as additional elements, which are important for the subunit interactions within the DegP hexamer. Interestingly, the PDZ1 domains are arranged differently in molecules A and B, whereas PDZ2 was only defined by electron density in molecule B. Although these differences may be influenced by crystal packing constraints, the pronounced mobility of the PDZ domains seems to be mechanistically important. The overall conformations of the protease (including the N terminus) and the PDZ1 domains are otherwise identical. When aligned separately, the root-mean-

squared deviation (r.m.s.d.) of the C α atoms of the protease domain was 0.54 Å, whereas the r.m.s.d. between the individual PDZ1 domains was 1.33 Å.

The DegP A and B hexamers are centred in the crystallographic unit cell at (0, 0, 0) and (2/3, 1/3, 1/4), respectively, and display crystallographic D3 symmetry. Crystal contacts are observed between the protease parts of A and B. The PDZ domains do not contribute to packing (see Supplementary Information Fig. 1). Notably, the observed hexamers represent two distinct states. Molecule A is a largely open structure with a wide lateral passage penetrating the entire oligomer, whereas molecule B corresponds to the closed form, in which a triangle-shaped 45 Å cavity containing the proteolytic sites is completely shielded from solvent (Fig. 1b). For both molecules A and B, the top and bottom of the DegP cage are constructed by the six protease domains, whereas the 12 PDZ domains generate the mobile side-walls. Because the axial pores of

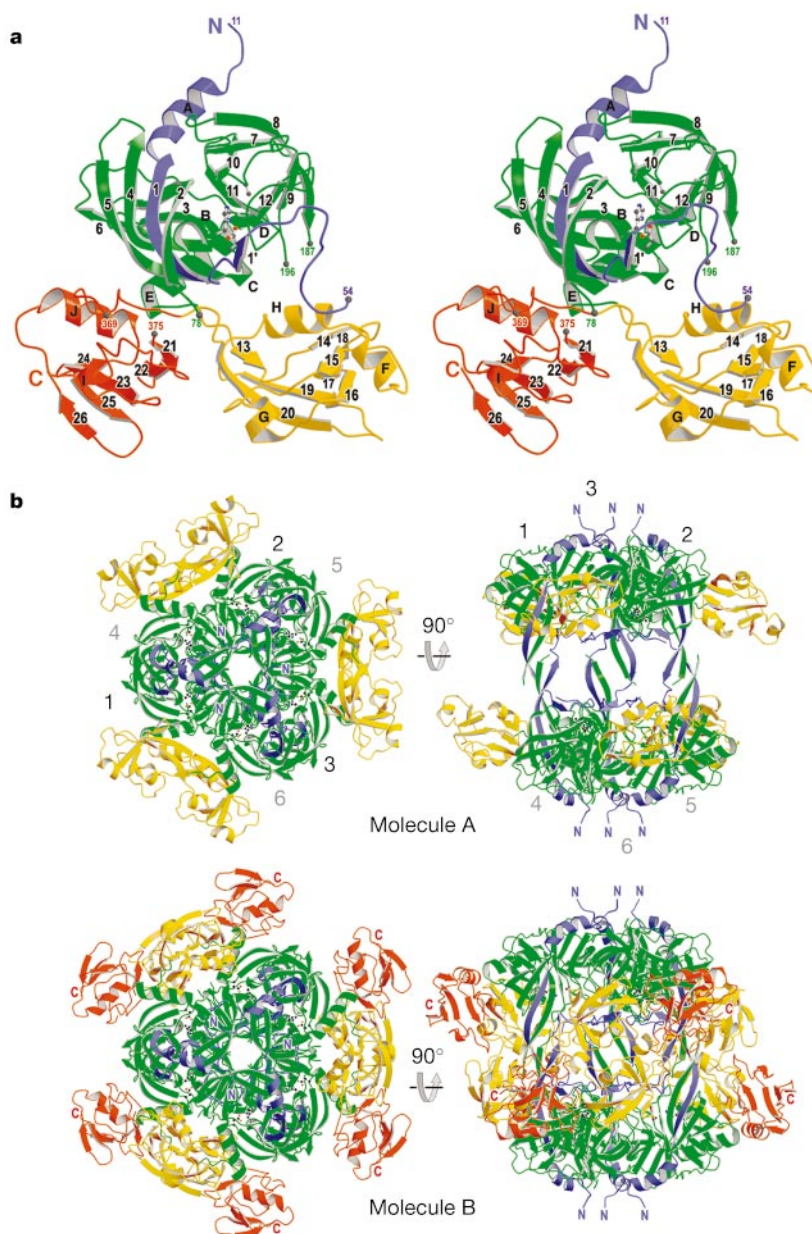


Figure 1 Structure of DegP. **a**, Stereo ribbon presentation of the monomer in which the individual domains are coloured differently. N terminal, purple; protease, green; PDZ1, yellow; PDZ2, red. Residues of the catalytic triad are shown in a ball-and-stick model. The nomenclature of secondary structure elements (helices by letter, strands by numbers), the

termini of the protein, and regions that were not defined by electron density are indicated. **b**, Top and side views of the DegP hexamer constructed by molecules A and B. Both hexamers are approximately equal in size, having a height of 105 Å and a diameter of 120 Å. The nomenclature of the individual monomers and their termini are given.

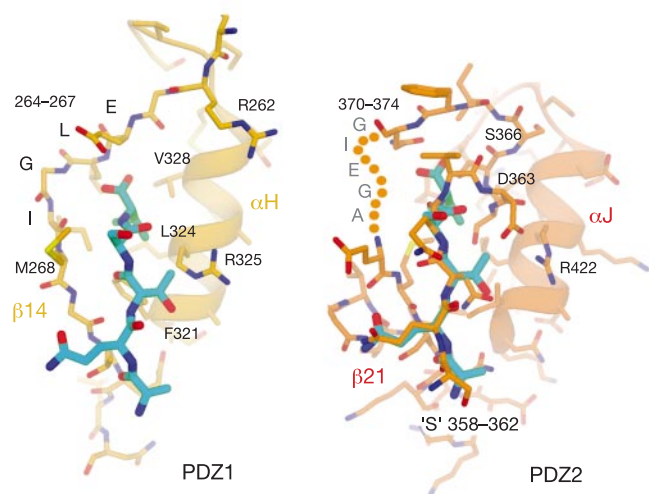


Figure 2 Peptide-binding sites of PDZ1 and PDZ2. The bordering β -strand, α -helix and carboxylate-binding loop are shown in a stick model and are coloured by atom type. After superposition with complexed PSD95, the peptide ligand (cyan) was modelled into the binding sites of both of the PDZ domains. Residues that may participate in substrate binding (PDZ1) and residues that anchor the substrate-like segment ('S', PDZ2) are labelled. The G-I-E-G-A motif of PDZ2 was not defined in the electron density, as indicated by the dashed coil.

the particle are blocked completely, the PDZ domains are the only gates allowing lateral access to the central cavity. This structural organization is markedly different from other cage-forming proteins, where substrates enter the central cavity through narrow axial or lateral pores^{12–17}. In particular, the PDZ1 domains interact with each other and should thus be the main gatekeepers of the inner chamber. The height of the cavity is determined by three molecular pillars, which are formed by enlarged loops of the protease domain. These pillars are also mainly responsible for the stability of the dynamic complex as seen in the open state.

The contacts between trimeric rings of the DegP hexamer arise almost exclusively from highly flexible structural elements. Thus, the DegP hexamer should be considered as a loosely bound dimer of tight trimers. Monomers associate into trimers mainly by hydrophobic interactions. The resulting trimeric rings interact through the polar 1–4 and 1–6 interfaces (Fig. 1b). In the 1–4 interface, the enlarged protease loops termed LA (connecting β 1 and β 2) are wound around each other and build the corner pillars of the DegP

cage. This molecular spacer is mainly stabilized by the two-stranded β -sheet 1'/2* (the asterisk denotes the participation of the neighbouring monomer). After reaching the opposite ceiling of the cavity, loop LA protrudes into the active site of its partner subunit. In the hexamer of molecule B, the 1–6 interface is formed by the interaction of PDZ1 and PDZ2 with their symmetry mates. The PDZ domains obtain a zipper-like arrangement, in which the PDZ1 domains are facing each other, whereas the PDZ2 domains are bound at their edges. Residues of this interface originate from the α F- β 15- α G segment (PDZ1) and from β 22- α I (PDZ2). In the open form of molecule A, PDZ1 tilts 70° away, thereby moving the α F- β 15- α G interaction clamp to the opposite side of the trimer–trimer interface (corresponding to a 30 Å movement), and breaking the 1–6 subunit interaction. The re-orientation of PDZ1 is achieved by a twist of the polypeptide backbone between residues Arg 262 and Gly 263.

In addition to their function in regulating access to the inner cavity, the PDZ domains of DegP should be involved in substrate binding. To identify the determinants of substrate specificity of PDZ1 and PDZ2, we aligned both structures with the peptide complex of the PDZ protein PSD95 (ref. 18), and modelled the bound ligand to both PDZ domains of DegP (Fig. 2). PDZ1 contains a deep binding cleft for substrate, which is mainly constructed by strand 14, its N-terminal loop (the 'carboxylate-binding loop'), and helix H. The carboxylate-binding loop is located in a highly positively charged region and is formed by an E-L-G-I motif, which is similar to the frequently observed G-L-G-F motif¹⁰. The side-chain of the strictly conserved Arg 262 is properly oriented to further fix the carboxylate group of the substrate. As this arginine is one of the hinge residues connecting protease and PDZ1, substrate binding to Arg 262 might trigger re-orientation of the PDZ domain. Binding specificity is conferred mainly by the specific configuration of the 0, –2 and –3 binding pockets¹⁹, where pocket 0 anchors the side chain of the C-terminal residue. In PDZ1, all pockets are built by residues that are mainly hydrophobic (Supplementary Information Fig. 2). One of the main differences to other PDZ domains is the flexibility of strand 14 and its associated carboxylate-binding loop, indicating the plasticity of the binding site. Thus PDZ1 seems to be well adapted to bind various stretches of hydrophobic peptide ligands.

The comparison between PSD95 and PDZ2 suggests a new mechanism of substrate binding and release. In PDZ2, the loop consisting of residues 355–369 adopts a conformation that mimics a bound substrate molecule (Fig. 2). Residues 358–362 are bound in extended conformation antiparallel to β -strand 21, and several specific interactions occur with helix J. This substrate-like segment

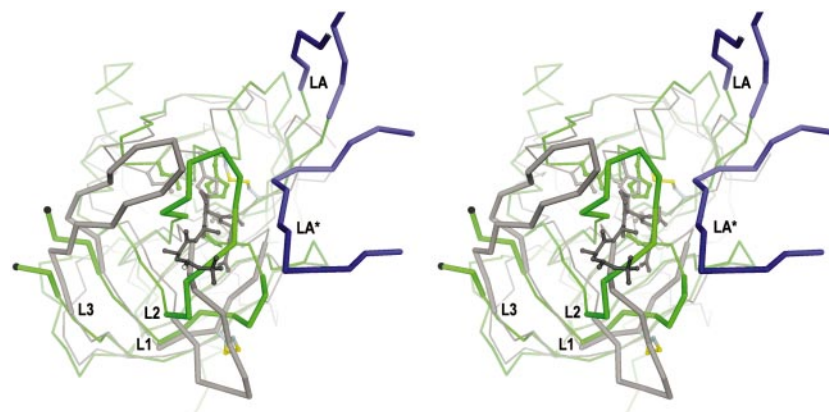


Figure 3 The protease domain. Structural alignment of the DegP protease domain (green and blue) with SGT (grey). In the stereo image, both backbones are shown as a C α trace, whereas the catalytic triads, the peptide bound to SGT (brown) and its disulphide bridges are drawn in ball-and-stick mode. Some of the mechanistically important loops (L1, L2, L3

and LA* (ref. 30)) are emphasized by thicker lines and are labelled. Note that loop LA* (blue) originates from the partner monomer. Loop L3 of DegP was only partially defined in the electron density.

is connected by the carboxylate-binding loop (comprising the G-I-E-G-A motif) to the remainder of PDZ2. The highly flexible carboxylate-binding loop might allow re-orientation of the substrate-like segment, and thereby the opening and closing of the peptide-binding site. As the occluded pocket was observed in the closed form of the DegP cage, displacement of a bound substrate by the loop 355–369 might be involved in the process of substrate translocation. As for PDZ1, some of the structural elements that might be involved in substrate binding, that is, the carboxylate-binding loop and strand 21, have considerably higher thermal motion factors than the rest of the domain, pointing again to a flexible binding site. Thus, both PDZ domains represent two forms of flexibility: a rigid-body movement generated by different twists of the hinge regions, and a local flexibility that should enable the accommodation of a wide range of substrates. Consequently, the PDZ domains are good candidates for the initial interaction with substrate.

The Dali algorithm was used to search for structural homologues of the protease domain. *Streptomyces griseus* trypsin (SGT), epidermolytic toxin, β -trypsin and thrombin were identified as the most similar structures in the database (Supplementary Information Table 1). Although the core of the protease domain is highly conserved, there are marked differences in the surface loops, which are important for the adjustment of the catalytic triad (Asp 105, His 135, Ser 210) and the specificity pocket S1 (Fig. 3). The enlarged loop LA protrudes into the active site of one monomer of the opposite trimeric ring, where it intimately interacts with loops L1 and L2. The resulting loop triad LA^{*}-L1-L2 completely blocks the entrance to the catalytic site. Loop L2 in particular is bent into a conformation that closes the active site, a feature that has not been reported previously. The resulting twist of the active-site loops impedes proper adjustment of the catalytic triad and formation of the oxyanion hole, as well as the S1 specificity pocket. Thus the protease domain of the DegP chaperone is present in an inactive state, in which substrate binding as well as catalysis is prevented. Furthermore, the widely conserved disulphide bond connecting loops L1 and L2 (Cys 191–Cys 220) is absent. This disulphide is believed to fix the relative orientations of both loops, thereby stabilizing the conformation of the S1 pocket. In DegP, loop LA^{*} fulfills a similar role and represents a flexible, non-covalent link between the two substrate-binding loops L1 and L2. To adopt the classical catalytic model of serine proteases, these loops have to undergo large conformational changes. Structural flexibility of these elements shown by their increased thermal motion factors might facilitate this process.

Escherichia coli DegP has the ability to stabilize a number of non-native proteins *in vivo* and *in vitro*^{2,20}. Possible binding sites for misfolded proteins are located within the inner cavity and are constructed entirely by residues of the protease domain. The solvent-accessible height of this chamber is 15 Å at its centre and increases to 18 Å near the outer entrance. Owing to these geometric constrictions, substrates must be partially unfolded to reach the active site (Fig. 4a). As in other chaperones of known structure, the DegP cavity is lined by hydrophobic residues, most of which are conserved. In both trimeric rings, three large hydrophobic grooves are organized around the central Gln 206/Arg 207 cluster and extend towards the PDZ1 domain (Fig. 4b). The hydrophobic grooves are mainly constructed by residues of loop LA and L2. A second potential binding site was observed on the internal side of the three pillars of the cavity. Furthermore, the hydrophobic binding sites of the PDZ1 domains are properly oriented to augment the number of potential binding patches. The alternating arrangement of polar and hydrophobic surfaces, both within one trimeric ring and between trimeric rings, should form the basis for binding exposed hydrophobic side chains and the peptide backbone atoms of substrates. Thus, the ceilings of the DegP cavity may represent docking platforms for partially denatured proteins, a construction reminiscent of a compactor (Fig. 4a). Both docking platforms are structurally flexible, as indicated by their backbone variations and by their high thermal motion factors. This plasticity should allow binding of diverse polypeptides.

Cage-forming proteases and chaperones can be energy dependent or energy independent. In the former group, ATPase activity is important for recognition of target proteins, their dissociation and unfolding, their translocation within the complex, and various gating mechanisms. The crystal structure indicates why these functions are not relevant for DegP. DegP preferably degrades substrates, which are, *per se*, partially unfolded and which might accumulate under extreme conditions^{4,21,22}. Alternatively, threading of substrate through the inner chamber could promote unfolding into an extended conformation. Removal of higher-order structural elements would allow the substrate to re-initiate folding after exit from DegP. Recruitment of PDZ domains for the gating mechanism should permit a direct coupling of substrate binding and translocation within the DegP particle. Accordingly, the PDZ domains may function as tentacular arms capturing substrates and transferring them into the inner cavity. By binding to the C terminus or a β -hairpin loop of a protein, the PDZ domains could properly position the substrate for threading it into the central cavity. After accessing this chamber, the fate of the unfolded protein depends on the interplay of loops LA, L1 and L2. □

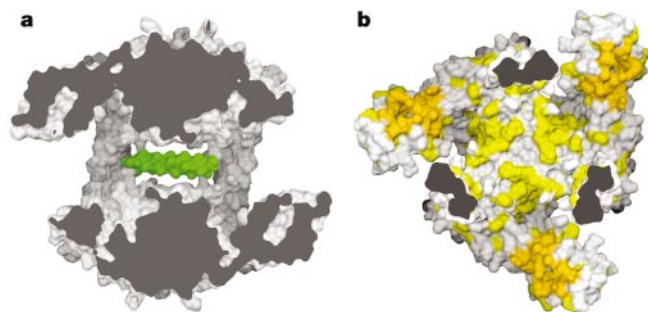


Figure 4 The central cavity. To demonstrate the structural properties of the inner cavity, cleaved presentations of molecule A were generated. Cut regions are shown in black. **a**, Surface representation of the internal tunnel (side view) illustrating its molecular sieve character. Access is restricted to single secondary structure elements as shown by the modelled polyaniline helix (green). **b**, Formation of the hydrophobic-binding patches within the cavity (top view). Hydrophobic residues of the protease domain are shown in yellow, and the non-polar peptide-binding groove of PDZ1 is in orange.

Methods

Crystallization and structure solution

Recombinant his-tagged protein was produced with *E. coli* strain CLC198 ($\Delta degP::Tn10$), as described previously². The three-step purification procedure included NiNTA, hydroxylapatite and gel filtration chromatography. Both S210A and S210A-SeMet proteolytically inactive mutants were crystallized at 18 °C using the vapour diffusion method and 10% isopropanol, 10% polyethylene glycol (PEG) 2000 monomethyl ether (MME), 0.1 M Tris/HCl, pH 8.5, as crystallization solution. Hexagonal crystal plates appeared within two weeks and belonged to space group P6₃22 with cell parameters $a = 121.3$ Å, $b = 121.3$ Å, $c = 237.2$ Å, corresponding to two monomers per asymmetric unit. For a two-day soak with a saturated Ta₂Br₁₄ solution, difference Patterson analyses yielded two tantalum sites, which then enabled us to identify four platinum sites in PtCl₄-soaked crystals. Subsequently, difference Fourier analyses yielded 21 out of the 28 theoretical selenium sites of SeMet crystals. However, the diffraction pattern of all DegP crystals was extremely anisotropic, restricting data collection to 4.5 Å resolution. This problem could be overcome by transferring crystals to 4 °C and exchanging the reservoir solution by 12% isopropanol, 50% PEG 2000 MME, 0.1 M Tris/HCl, pH 8.5. After one week, the equilibrated crystals could be flash frozen directly in liquid nitrogen. This treatment resulted in an isotropic diffraction pattern to 2.8 Å resolution. A complete multiwavelength anomalous diffraction experiment was performed with the SeMet crystals at the high-brilliance beamline ID14-4 at the European Radiation Synchrotron Facility (ERSF; Supplementary Information Table 2).

Model building and refinement

Energy-restrained crystallographic refinement was carried out with maximum likelihood algorithms implemented in CNS²³, using the protein parameters of ref. 24. Bulk solvent, overall anisotropic *B*-factor corrections and non-crystallographic restraints were introduced depending on the behaviour of the *R*_{free} index. Refinement proceeded in several cycles, which were interrupted for manual rebuilding with the program O²⁵. After the addition of solvent molecules, the refinement converged at an *R*-factor of 21.8% (*R*_{free} = 27.5%). We prepared all graphical presentations using the programs MOLSCRIPT²⁶, RASTER3D²⁷, SETOR²⁸, GRASP²⁹ and DINO (<http://www.dino3d.org>).

Received 21 November 2001; accepted 14 February 2002.

- Wickner, S., Maurizi, M. R. & Gottesman, S. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**, 1888–1893 (1999).
- Spiess, C., Beil, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**, 339–347 (1999).
- Gottesman, S., Wickner, S. & Maurizi, M. R. Protein quality control: triage by chaperones and proteases. *Genes Dev.* **11**, 815–823 (1997).
- Lipinska, B., Zylicz, M. & Georgopoulos, C. The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. *J. Bacteriol.* **172**, 1791–1797 (1990).
- Pallen, M. J. & Wren, B. W. The HtrA family of serine proteases. *Mol. Microbiol.* **26**, 209–221 (1997).
- Zumbrunn, J. & Trueb, B. Primary structure of a putative serine protease specific for IGF-binding proteins. *FEBS Lett.* **398**, 187–192 (1996).
- Gray, C. W. *et al.* Characterization of human HtrA2, a novel serine protease involved in the mammalian cellular stress response. *Eur. J. Biochem.* **267**, 5699–5710 (2000).
- Suzuki, Y. *et al.* A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death. *Mol. Cell* **8**, 613–621 (2001).
- Wootton, J. C. & Drummond, M. H. The Q-Linker—a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.* **2**, 535–543 (1989).
- Cabral, J. H. M. *et al.* Crystal structure of a PDZ domain. *Nature* **382**, 649–652 (1996).
- Liao, D. I., Qian, J., Chisholm, D. A., Jordan, D. B. & Diner, B. A. Crystal structures of the photosystem II D1 C-terminal processing protease. *Nature Struct. Biol.* **7**, 749–753 (2000).
- Lowe, J. *et al.* Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539 (1995).
- Bochtler, M., Ditzel, L., Groll, M. & Huber, R. Crystal structure of heat shock locus V (HslV) from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **94**, 6070–6074 (1997).
- Wang, J. M., Hartling, J. A. & Flanagan, J. M. The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cell* **91**, 447–456 (1997).
- Larsen, C. N. & Finley, D. Protein translocation channels in the proteasome and other proteases. *Cell* **91**, 431–434 (1997).
- Groll, M. *et al.* A gated channel into the proteasome core particle. *Nature Struct. Biol.* **7**, 1062–1067 (2000).
- Brandstetter, H., Kim, J. S., Groll, M. & Huber, R. Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature* **414**, 466–470 (2001).
- Doyle, D. A. *et al.* Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85**, 1067–1076 (1996).
- Songyang, Z. *et al.* Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **275**, 73–77 (1997).
- Misra, R., Castillockeller, M. & Deng, M. Overexpression of protease-deficient DegP(S210A) rescues

the lethal phenotype of *Escherichia coli* OmpF assembly mutants in a degP background. *J. Bacteriol.* **182**, 4882–4888 (2000).

- Swamy, K. H. S., Chung, C. H. & Goldberg, A. L. Isolation and characterization of protease Do from *Escherichia coli*, a large serine protease containing multiple subunits. *Arch. Biochem. Biophys.* **224**, 543–554 (1983).
- Strauch, K. L., Johnson, K. & Beckwith, J. Characterization of *degP*, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature. *J. Bacteriol.* **171**, 2689–2696 (1989).
- Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr. A* **47**, 392–400 (1991).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Kraulis, P. J. Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Bacon, D. J. in *Macromolecular Crystallography* Part B 505–524 (Methods in Enzymology, Vol. 277, Academic, New York, 1997).
- Evans, S. V. Setor—hardware-lighted 3-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134–138 (1993).
- Nicholls, A., Bharadwaj, R. & Honig, B. Grasp—graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166–A166 (1993).
- Perona, J. J. & Craik, C. S. Structural basis of substrate specificity in the serine proteases. *Protein Sci.* **4**, 337–360 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com>).

Acknowledgements

We thank C. Spiess for providing cells for SeMet expression, and G. P. Bourenkov and J. McCarthy for assistance during data collection at beamlines BW6 (DESY, Hamburg) and ID14-4 (ESRF, Grenoble), respectively. We thank R. A. John and W. Bode for critical reading of the manuscript. Financial support by EU ERB (M.G.-F.), DFG (R.H.) and Aventis (T.C.) is acknowledged.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.C. (e-mail: clausen@biochem.mpg.de). The atomic coordinates and structure factor amplitudes of the DegP S210A crystal structure have been deposited at the Protein Data Bank under entry code 1KJ9.

Tools of the proteomics trade

Proteins handled *en masse* and individual proteins isolated and identified.

Evolution 200

Presearch www.presearch.co.uk

HPLC workstations for proteomics

The Evolution Nano LC Workstation can run flows down to 1 µl per min and can also be set up to provide Nano LC flows using a pre-column split. The makers claim the 7-port valve system provides the lowest mixing volume of any Nano LC system available, reducing run times and increasing productivity. The system has options for advanced automation including on-line sample trapping, SDS removal and fraction collection direct to MALDI targets. Evolution 200 can be interfaced to all leading ion-trap and quadrupole TQF mass spectrometers.

AntibodyArrays

Hypromatrix www.hypromatrix.com

Zoom in on interesting proteins

AntibodyArrays are designed to provide rapid and cost-effective identification of protein-protein interactions, post-translational modification and protein expression. AntibodyArray is a membrane on which antibodies to a variety of protein ligands are immobilized. When cell lysates are applied and incubated, proteins bind to their respective antibodies. The protein of interest is detected by exposing the membrane to an antibody against that protein. Proteins with which the target protein interacts are identified from their position on the membrane.

Pro ID

Applied Biosystems

www.appliedbiosystems.com

ID check for proteins

The Pro ID software program was developed for automated protein identification in complex biological samples when used for analysis of tandem MS proteomics data. The software can be used on the API QSTAR Pulsar Hybrid LC/MS/MS system. Pro ID performs error and modification-tolerant searches using a zone modification feature that can identify modifications in the absence of prior knowledge about them.

ProteoLink

Promega www.promega.com

Find the missing link

The ProteoLink system links phenotype with genotype using a plasmid-based human adult brain cDNA library expressed through Promega's TNT reticulocyte-based coupled transcription/translation system. ProteoLink

can be used for substrate identification, enzymatic activity and interaction screens. Screens can be performed for protein-protein, protein-DNA and protein-lipid interactions. Advantages over other screening assays are said to include accurate protein folding, post-translational modifications, options to include additional activities, and high throughput.

ProteomeWorkssystem.com

Bio-Rad www.bio-rad.com

Proteomics on the web

Promising comprehensive practical information, Bio-Rad describes its new website as a resource for researchers working in all aspects of protein discovery. The site features the ProteomeWorks system and provides protocols for a range of techniques involving sample preparation. The resource section includes links to bioinformatics tools, protein databases and research laboratories.

Relay 96

Invitrogen www.invitrogen.com

High-throughput protein solubility screening

The Relay 96 protein screen from Invitrogen allows protein solubility screening of up to 96

samples in 30 minutes. Researchers can load samples directly, without cell harvesting, then sequentially extract soluble and insoluble fractions from the same sample. A solid-phase lysis matrix prevents bacterial chromosomal DNA contamination and eliminates sample viscosity.

Data Mining Components

Numerical Algorithms Group www.nag.co.uk

Dig deep into your data

Data Mining Components is a software collection containing numerical algorithms designed to aid data mining modelling processes in bioinformatics and other fields. The algorithms are used to identify patterns and trends within data and can be applied throughout the modelling process, from data cleaning through data transformation to model building. The software can be set to 'essential inputs only' mode for rapid prototyping or 'full interface' mode for greater control. It can be integrated into new or existing data mining applications and is available in Windows, Linux or Sun Solaris versions.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayouni (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy: Nevin Bayouni (4978)

Scandinavia: Silje Opstrup (4994)

Spain/ Portugal: Leonie Welss
(4954)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel +33 (0) 1 43 20 16 51
Fax +33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel +49 89 54 90 57 11/-2
Fax +49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakakamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com

Japan Manager: Kate Cowan

naturejobs

Chemists buck the trend

Although the US economy weakened last year, job prospects for newly graduated chemists did not, according to a recent report by the American Chemical Society (ACS). The society's recent salary survey shows that for 2001 master's and PhD graduates, salaries rose by 6–8% compared with the previous year. The median master's recipient salary last year was \$48,000, a \$3,000 increase over the previous year. New PhDs fared even better, earning on average \$70,000, up \$5,000 from the class of 2000. And unemployment remained unchanged from 2000, at 3% for new PhD chemists and 5% for freshly minted master's.

The big question remains whether or not these numbers provide a true picture of the present, as they were collected up to October 2001. The economy continued to weaken after that point and, the survey notes, more chemists are going into pharmaceuticals — an area that is showing signs of caution in terms of recruitment this year.

Weak economy or not, chemists have something working in their favour: fewer are being produced in the United States. So unless demand dips sharply, work should be reasonably easy to find — although not necessarily in their sector of choice.

Academia seems a relatively stagnant market for chemists, as it features significantly lower pay than the private sector. Also, traditional industrial chemical manufacturing is fading, but pharmaceuticals and analytical research firms have seen a steady growth in their need for chemists.

A good bet would be that demand will increase again in the second half of the year, but salaries will not go up quite as much. The question of whether or not pharmaceuticals will consolidate further remains one of the biggest unknown factors.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS & STUDENTS

Bringing the brains
back home p5

MOVERS

Canberra gets new vice-
chancellor; one of Dolly's
creators heads for Singapore;
high-energy physicist takes up
the SLAC; and more Back page

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Easing the journey back home

Thomas Mayer:
A repatriation
award contributed
significantly to my
decision to go back to
Germany.



Each year, many young European scientists start postdoctoral work in the United States. But on completion, they can face difficulties in securing a position back home. Repatriated scientists, even when highly qualified, often lose out to better-connected internal candidates if they don't have strong backing within the department.

Last year, Spain launched a programme that has already brought a sizeable number of its scientists back (see *Nature* **413**, 556; 2001). Among other countries that send a large number of scientists abroad with stipends are the United Kingdom, Germany and France. Granting agencies in each country have come up with different approaches to ease the transition.

Frank Gannon, executive director of the European Molecular Biology Organization (EMBO), sees these programmes as a start, and a "component of lowering barriers". Improving each country's quality of science is a long-term way to retain scientific talent, he adds.

Enric Banda, secretary general of the European Science Foundation, agrees that the scientific environment is most important. He would like to see more independent positions that let young scientists run their own projects. "Europe has certainly identified this issue," he says, "but actions are slow, let alone results."

WELLCOME HOME

At the start of this year, Britain's Wellcome Trust awarded the first of its International Research Fellowships (IRFs), which provide up to four years of support. After two to three years outside Britain or Ireland, holders receive up to two more years' repatriation support.

Bruce Turnbull, a chemist at the University of Leeds, received the IRF's precursor award, the International Prize Travelling Research Fellowship (IPTRF). This provided funding for two years abroad plus an additional year after the return. Turnbull says that many of his European colleagues at the University of California, Los Angeles, had problems arranging jobs back home. Turnbull would have considered staying in the United States, but the repatriation funding made going home an "easy decision", he says.

Postdocs Richard Wade-Martins and Eric Miska, two current IPTRF holders in Boston, agree that this type of fellowship provides generous support, but they

feel that two years abroad may not be enough to complete a project and get the results published. They both see the IRF's extra year as a big improvement.

GERMAN APPROACH

Germany's main grant agency, the DFG, aids postdocs with its Emmy Noether Programme. Phase one gives two years of support for a postdoctoral research project abroad; the second phase adds funding to establish a small research group back in Germany.

Cell biologist Thomas Mayer left Harvard last year to set up a lab in Germany. Mayer, now at the Max Planck Institute of Biochemistry near Munich, is glad that he did not have to seek grants to get started. Enjoying a level of independence that is unusual in his country, he says that the Emmy Noether award "contributed significantly" to his decision to return to Germany.

Emmy Noether fellow Burckhard Seelig is slightly concerned that two years may not be enough time to finish a project. So the German biochemist, like many IPTRF recipients, secured funding while at Massachusetts General Hospital and will seek an extension before going home.

CNRS EFFORT

The French national research foundation CNRS helps about 3,000 French scientists in the United States and, together with the French embassy, organizes an annual careers fair called Forum USA. According to the head of the Washington office, Dominique Martin-Rovet, the CNRS fills 15% of its vacant positions with applicants working in the United States.

Cyril Delattre, a chemical engineer at the Massachusetts Institute of Technology, attended Forum USA to get information about opportunities in France's private sector. "It's very helpful, you don't have to fly to France to talk to all the human resources managers," says Delattre. But he adds that applying for an academic post at the CNRS is more difficult. Before submitting a research proposal it is vital to visit potential labs and secure the support of the principal investigator. According to Delattre, this is no small feat if you "feel disconnected" by distance.

Jan Schmollinger is a graduate student in Boston.

Web links

Wellcome Trust
♦ www.wellcome.ac.uk
DFG
♦ www.dfg.de
CNRS
♦ www.cnrs.fr
Forum USA
♦ www.forumusa.org

Transitions

X Martin Hollenhorst will take up his new post as chief financial officer for Heidelberg-based LION Bioscience in April. Hollenhorst comes to LION after spending some time in Washington DC as a freelance management consultant. He will replace **Klaus Sprockamp**, who died in the 11 September attack on the World Trade Center.

X Damian Marron was this month appointed senior director, business development, for NicOx, a drug firm based in Sophia Antipolis, France. Marron is a former head of European business development for 3M Pharmaceuticals, which he joined in 1997.

X Last month Ardana Bioscience, an Edinburgh-based biotech company focused on women's reproductive health, appointed **Finn Larsen**, previously medical director at Columbia Laboratories, Paris, France, as its clinical director.

X Last month **Lester Crawford** was named deputy commissioner of the US Food and Drug Administration. Crawford most recently served as head of the Center for Food and Nutrition Policy at Virginia Tech in Blacksburg.

X Julia King, director of marine engineering and technology at UK-based company Rolls-Royce, will this September succeed **Alun Jones** as chief executive of the Institute of Physics.

X Michael Wokasch, president of Vertex Pharmaceuticals subsidiary PanVera, had his role expanded this month. In addition to his post as PanVera, he will now serve as president of another Vertex subsidiary, Aurora Biosciences, where he replaces **Harry Styli**.

HIGH-ENERGY PHYSICS

After 14 years at Cornell University in Ithaca, high-energy physicist **Persis Drell** is moving to the Stanford Linear Accelerator Center (SLAC) in Menlo Park, California. She steps into her new role as associate director of SLAC's research division next month. She replaces **Steve Williams**, who had been acting associate director since September 2000.

Drell says she will miss Ithaca, but she feels that SLAC's larger size and broader programme — which she believes could help scientists to push the boundaries of the standard model — will be adequate compensation.

One of the attractions of her new post is the combination of additional management responsibilities and a continued freedom to be involved in experiments. "This lets me stretch my administrative wings but keeps me close to the science," Drell says.

Drell's decision to move was made easier by the fact that Stanford has also hired her husband, **Jim Welch**, who is an accelerator physicist. This solved the "two-body problem" that can limit the options for some academics, she says.

BIOCHEMISTRY

Roger Dean, former director of the Heart Research Institute in Sydney, last month assumed the position of vice-chancellor at the University of Canberra. Dean describes himself as a "biochemist and musician" — as well as his scientific interests, he is a keen composer involved in musical improvisation. He says that his new position will allow him to use "both sides of my brain", as he will be involved in both science and the humanities.

Dean plans to expand the teaching and research efforts at the university, and extend both its commercial and community outreach.

BIOTECHNOLOGY

Alan Colman, one of the scientists who helped to clone Dolly the sheep in 1996, is to leave his post as research director of Scottish firm PPL Therapeutics to become chief scientific officer at ES Cell International in Singapore. Colman is moving primarily because he believes that Singapore will offer a better investment climate for stem-cell research compared with Britain.

Australia's Monash University, a world leader in stem-cell technology, was one of the founding institutions behind ES Cell International. The company's location allows it to benefit from light regulation



Persis Drell



Roger Dean



Elaine Fuchs



Noëlle Lenoir

and a share of a £400-million (US\$570-million) biomedicine investment fund, which was set up by the Singapore government in 2000.

The announcement of Colman's move comes less than a year after **Roger Pedersen** decided to leave a faculty position at the University of California, San Francisco, for a similar post at the University of Cambridge because of Britain's more permissive policies towards stem-cell research (see *Nature* **412**, 262; 2001).

CELL BIOLOGY

The "20-year-itch" has helped to convince **Elaine Fuchs**, past president of the American Society for Cell Biology, to leave the University of Chicago and head for Rockefeller University to lead the Mammalian Cell Biology and Development Laboratory. Fuchs and her friend **Susan Lindquist**, who last year left Chicago to direct the Whitehead Institute in Cambridge, Massachusetts (see *Movers*, *Naturejobs* 30 August 2001), had both done some soul-searching and decided to shake up their careers a little.

Although she enjoyed her 20 years at Chicago, the past few years left Fuchs feeling that things were getting a little too familiar. Rockefeller will give her "a new environment with new colleagues taking different approaches to different problems", she says. She is especially interested in interacting with neuroscientists and structural biologists to develop her research in the biology of the skin.

LAW

Noëlle Lenoir, former Justice of the French Constitutional Supreme Court, is enjoying her new position as a biotech lawyer with the firm Herbert Smith in Paris, which she took up this winter. "In France it's not so easy to go from the public sector to the private sector," she says. But she notes that working in the public sector "opened a lot of doors" for her.

Apart from her Supreme Court position, Lenoir served on bioethics panels for both the United Nations and the European Union. Doors are now starting to open for her in the French biotech sector, she says, which is beginning to expand as a result of increased investment.

Lenoir serves the law firm as a specialist in biotech patents and pharmaceutical law. At the same time she is a visiting professor at University College London and Yale University in Connecticut. "My life is quite an exciting challenge for a former Justice," she says.

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:
▶ naturejobseditor
@naturedc.com.